数学辞海

MATHEMATICS DICTIONARY

第五卷

山 西 教 育 出 版 社
东 南 大 学 出 版 社
中国科学技术出版社

山西省人民政府资助出版

如果你有数学問題,
而不好意思發問,
你可从本書获得解答.

一九八八年八月書祝
数学辭海成功

陳省身



学习、研究、运用、发展数学，让中国数学赶上国际先进水平，促进社会主义現代化建设

吴大任

数学的进步和完美与国家的繁荣和昌隆是紧密相连的

杨乐 八六.七.

長江納衆水

百折不回頭

碧海能容物

滔滔向海流

數學辭海出版紀念

李國平并書





《数学辞海》总编辑委员会

《数学辞海》第五卷编辑委员会

《数学辞海》第五卷各分支学科编辑委员会

序

当我们向着日益临近的21世纪走去的时候，一部巨著——《数学辞海》将要面世了。这是我国200余所高等院校、科研机构，数以千计的数学家、数学工作者共同劳动的结晶，是一件影响深远的大事。

还是在本世纪同上一世纪交接的1900年，希尔伯特就以23个数学问题作为送旧迎新的礼物，高瞻远瞩地指引着20世纪数学发展的若干重要进程。如今，20世纪的帷幕行将落下，我们惊喜地看到，在这百年间，数学已经发生了多么巨大的变化！人们对数学的认识更深刻了，数学的分支更多了，数学的广度和深度，都远远超出了本世纪初的预料。异军突起的新科学和新技术，诸如计算机科学、航天技术、生命工程、数字通信以及新能源的开发、新材料的应用等，无不需要数学，社会科学和人文科学的经济、教育、语言、考古等领域，也开始与数学结下不解之缘。所有这些学科在向数学索取的同时，也都在某一方面推动和改变着数学。数学已经发展成为内涵广泛的数学科学。数学是大自然的语言，又是人类社会生活中各种关系的高度概括。数学在现实世界中获取模型，扩大了自己的外延，同时展现了新的内涵、新的抽象。如果说古希腊和古代中国的数学只是涓涓细流，那么，今天之数学已经汇成了波流浸灌的长江大河。

一个人可以学贯中西，但无法懂得现代数学的方方面面，而社会变革的进程和新技术的浪潮却迫使人们学习和应用更多的数学。解决这个矛盾的办法之一，自然是编纂一部大型的数学工具书。《数学辞海》正是在这样的时代需求背景之下应运而生的。有了这种巨大的推动力量，它才能克服种种艰难曲折，从第一页稿纸，发展成为我们所见到的这部别具一格的鸿篇巨制。

为什么这本书能使作者们激动，愿意竭诚为之构筑，又能吸引读者，使之企足而待呢？这是由于数学自身的地位和价值，它在实践中的巨大作用和自身的美。

数学首先是人们生活和生产的工具。马克思非常赞同康德的这样一句话：“一门科学，只有当它成功地应用数学时，才算达到了完善的地步。”事实上，数学被使用的程度，往往反映了一个国家、一个民族的科学进步和经济发展水平。很难设想，在一个低技术的国家，会产生高深的数学，而高技术的社会形态，必有与之相适应的数学水平。毫无疑问，在科学技术飞速发展的当今世界，对数学的需求将与日俱增。

其次，数学又是一种文化形态。古往今来，人们在数学这块沃土上耕耘，收获了许多硕果。这些美好的硕果，本身就是一首首动人的诗篇，闪耀着智慧的光芒。一般人都会欣赏艺术，然而，当一个人只要具有基础的数学知识，同样可以对一道经典的数学名题和某个优美的解法叹为观止。人们还概括了大量实际模型的抽象数学，通过形式推演，以得出

系统的理论，再应用到更广泛的总体上去。数学的这种以简驭繁的本领决定于它的高度概括性。正是由于概括，数学形成了包含各个学科的优美结构。数学的发展推动了自然结构观的发展，它有力地带动了其他学科，大大加速了人类文明史的进程。

数学的作用，还在于它有着独特的培育人的功能。数学是每个人必须学习的基础学科。从小学到中学，数学的学时最多，除了因为数学是一切科学的基础和工具之外，更因为数学有着独特的思维教育和智力开发的作用。数学的高度抽象、遵从逻辑规则和不断创新的特征，集中而突出地表现了人类思维的概括性、逻辑性和探索性。所以，学习数学对人才的培养是一种基本的思维训练，被称为“思想的体操”。

为了全面地反映古今中外的数学成果、体现数学的多种功能，本书既兼顾传统数学和现代数学，又兼顾抽象的基础数学和具体的应用数学。考虑到广大数学教育工作者的需求，本书还将初等数学和高等数学相对地进行了划分，并按习惯将某些分支学科集中列卷，此外还编纂了包含数学史与数学教育等分支学科的专卷，也系统地介绍了中国的古算。这样编纂的《数学辞海》将会充分地显示数学的工具意义、文化意义和教育意义。愿这部国人自编的《数学辞海》既能为国家经济建设服务，又能在民族文化建设中起到应有的作用。

《数学辞海》是改革开放的产物，又将为改革开放服务。人们或许没有想到，这部巨著竟出自民办的编写组织。编纂者来自大江南北、长城内外、海峡两岸，在历时 10 余年间，数百所大专院校、科研机构的千余名专家学者日夜辛劳、自愿奉献，在《数学辞海》中编织着自己的理想和愿望。社会各界积极赞助，有识之士慷慨捐赠，海外同胞亦纷纷来电来函表示支持，用他们的心意渲染着文化建设的宏图。在这个民办写作团体中，人们互相信任、互相支持、互相勉励，充满了成就事业的认同感、紧迫感。这一写作经验也清楚地说明：在共同的愿望下，民办科研也是一条坦途。这是《数学辞海》编写过程中给我们的一个十分有益的启示。

像一切事物一样，《数学辞海》还不会达到绝对完善的境界。相反，这部反映浩如烟海的数学知识，动员了如此巨大力量而编纂的大型著作，首次面世时，一定会有许多不足之处，例如整体结构、条目收集、词义诠释、词类归属等，都还会有需要进一步推敲、商榷的地方。数学是极为严谨的科学，《数学辞海》必将在众多专家的严谨尺度之下，逐版改进。我们今天为之高兴的是，将来可能成为传世之作的《数学辞海》已有了良好的雏形，我们准备将它作为迎接新世纪的礼物，奉献给关心它、需要它的广大读者。

程民德

1998 年 6 月

凡　　例

一、编　排

1. 全书包括数学科学的100余个分支学科或专题项目，按照从初等数学到高等数学，从古典数学到现代数学，从理论数学到应用数学的原则，将整个数学科学划分为6卷编辑出版(参见《数学辞海》六卷本内容划分方案)。

2. 各卷正文均按数学知识的结构体系编排，同一分支学科(或同一专题项目)的条目相对集中，一般按知识本身的结构、层次、逻辑等关系确定其先后顺序，而数学史部分，如数学家、数学名著、数学期刊、数学团体等，则分别按其出生、出版、创刊、成立的年代先后顺序编排。

3. 各卷目录标题分为三级：一级标题为一个分支学科或一个知识门类。一级标题之下，则按知识构成设若干个二级标题。例如，第一卷中的“数学分析”为一级标题，下设六个二级标题——“实数理论”、“变量与函数”、“极限理论”、“微分学”、“积分学”和“无穷级数”；又如，第六卷中的“中国数学史”为一级标题，下设四个二级标题——“中国古代数学史”、“中国古代数学著作”、“中国古算名词术语”和“中国数学家”。三级标题为具体条目名称。

4. 同一卷中不同分支学科之间的内容有重复时，其知识主题一般地只在一处设条目；不同卷中的学科内容有重复时，其知识主题在各相关部分均设条目，但在释文内容上各有侧重。

二、条　目

1. 条目的标题一般为一个词，如“圆”、“群”、“环”、“函数”、“矩阵”、“向量”、“方程”等，也有的是一个词组，如“勾股定理”、“常微分方程的通解”、“哥德巴赫猜想”、“希尔伯特第6问题”等。

2. 条目设立的条件：1) 独立的知识主题或已形成的固定概念；2) 能够应用准确的、人们习惯和易于理解的词标引；3) 便于读者快速查阅。

3. 条目的分类：条目按其释文的长短分为五类：特长条目(3000字左右)、长条目(1000—3000字)、中条目(300—1000字)、短条目(300字以内)和参见条目。

4. 本书所收的数学名词术语，力求与“全国自然科学名词审定委员会”公布的《数学名词》(科学出版社，1993)保持一致。外国人名的中文译名，力求与《中国大百科全书·数学卷》和梁宗巨主编的《数学家传略词典》(山东教育出版社，1989)中的译名保持一致。未出现在上述著作中的外国人名的中文译名，则采用数学界的约定译名或用习惯译法译出的译名。

三、释　文

1. 本书条目的释文，以文字叙述为主，并采用规范化的现代汉语，力求科学、准确、简明、通俗，杜绝教材式语言和口语，释文开头不再重复条目的标题。

2. 释文开头一般要求破题，然后给出严格的数学定义，并尽量阐明该条目内容的历史沿革及其在本分支学科或知识门类中的地位、作用、发展趋势等，以增强释文的科学性和可读性。

3. 一词多义的释文，用①…②…③…分项叙述，某个条目的释文需由其他条目释文补充说明的，采

用“参见”的方式，被参见的条目标题需加引号，条目标题前加“参见”字样，并置于圆括号之内。

4. 对常见的异名同义词，只给出一种条目标题的释文，其余异名条目亦列入正文，但不再写释文，给出释文的条目标题加引号，条目标题前加“即”字样。例如：矢量(vector)即“向量”;全纯函数(holomorphic function)即“解析函数”;正则函数(regular function)即“解析函数”。

5. 每一个条目标题后，一般在圆括号内标注有对应的英文。凡外国人名(日本人除外)在条目的释文中第一次出现时，在其中文译名后的圆括号内标注有相应的外文原名的姓和名的首字母，并用逗号隔开。例如，欧几里得(Euclid)、牛顿(Newton,I.)、高斯(Gauss,C. F.)。同一外国人名在条目的释文中第二次出现时，不再标注外文。在日本人名、中国人名、中国古代数学史、中国古代数学著作、中国古算名词术语等条目标题后，一般在圆括号内标注汉语拼音。

6. 如果条目乙的基本定义已经完全包括在条目甲的释文之中，那么条目乙只作为“参见条目”保留，所参见的条目标题需加引号，并在条目标题前加“见”字样，而释文不再重复。例如，在条目“线性变换”的释文中，已给出“单位变换”、“恒等变换”和“零变换”的定义，则上述三个条目就作为“参见条目”予以保留，并分别表示为：单位变换(unit transformation)见“线性变换”；恒等变换(identity transformation)见“线性变换”；零变换(null transformation)见“线性变换”。

四、索　引

1. 本书每一卷正文之后，均附有三种索引，即条目笔画索引、条目音序索引和条目西文索引。索引中条目标题后面的数字，表示该条目在正文中的页码。

2. 在条目笔画索引中，以汉字起首的条目标题按第一字的笔画由少到多的顺序排列，若笔画数相同，则按 一(横)、丨(竖)、丿(撇)、丶(点)、 乛(折)五种笔形顺序排列，其中，㇀(提)归为一(横)，亅(竖钩)归为丨(竖)，㇏(捺)归为丶(点)，各种笔形带钩或曲折的笔画(除竖钩“亅”外)归为 乛(折)。第一字相同的，则按第二个字的笔画数和起笔笔形的顺序排列，依次类推。

3. 在条目音序索引中，以汉字起首的条目标题按第一字的汉语拼音字母顺序排列，若第一字的声母、韵母相同，则按声调的阴平、阳平、上声、去声顺序排列。第一字相同的，则按第二个字的汉语拼音字母顺序排列，多音字按不同的拼音字母顺序排列，依此类推。

4. 在条目笔画索引和条目音序索引中，凡第一字为西文字母、数学符号、罗马数字和阿拉伯数字起首的条目标题，一律排在两种索引的最后。西文字母起首的条目标题分别按其字母的花体、大写、小写及字母本身的先后顺序排列；数学符号起首的条目标题按知识结构顺序排列；数字起首的条目标题按由小到大的顺序排列。若起首的字母、符号及数字相同时，仍按其后汉字的笔画或音序排列。

5. 在条目西文索引中，按条目标题的起首西文字母顺序排列；条目标题的西文缩写，按一个词排列。凡以数学符号、罗马数字和阿拉伯数字起首的条目标题，一律排在条目西文索引的最后。数学符号起首的条目标题按知识结构顺序排列；数字起首的条目标题按由小到大的顺序排列。若条目标题起首的字母、符号、数字相同时，则按第二个字母等的顺序排列，余此类推。没有西文译名的条目，未收进条目西文索引。

6. 在各卷索引之后，还附有本卷涉及到的中外人名译名对照表，以供读者查阅。

目　录

《数学辞海》六卷本内容划分方案

第一卷

数学
算术
初等代数
平面几何
平面三角
立体几何
球面几何
平面解析几何
空间解析几何
初等数论
高等代数
高等几何
数学分析
集合论
形式逻辑
布尔代数
概率论与统计学初步
数学符号表

第二卷

数学
组合学
线性与多重线性代数
群及其推广
李群与李代数
环与代数
模与同调代数
序与格
范畴论与代数 K 理论
域论与伽罗瓦理论
数论
代数几何
微分几何学
凸集几何与距离几何
一般拓扑学
代数拓扑学与流形拓扑学
奇点理论与突变理论
数学符号表

第三卷

数学
实变函数论
复变函数论
多复变函数论
测度论
泛函分析
变分法
函数逼近论
调和分析
流形上的分析
位势论
凸分析
非标准分析
小波分析
分形几何
常微分方程
偏微分方程
积分方程
动力系统
特殊函数
数学符号表

第四卷

数学
数学基础
数理逻辑
计算数学
概率论
随机过程
统计学
经济数学
生物数学
数学物理与理论物理
模糊数学
数学符号表

第五卷

数学
运筹学
系统理论
控制理论
通信与信息理论
画法几何与工程图学
计算机科学
数理语言学
力学
天文学
测绘学
数学符号表

第六卷

数学
中国数学史
外国数学史
数学符号史
数学团体与研究机构
数学竞赛与数学奖
数学期刊
数学教育
数学哲学
数学名题与猜想
珠算
数学发展史年表

第五卷条目目录

说明：该目录由本卷所属各分支学科或专题项目的全部条目(包括给出释文的条目及其参见条目)组成，按知识结构顺序编排，即按条目在正文中出现的先后顺序排列.

运 筹 学

最 优 化

线 性 规 划

对偶线性规划

特殊的线性规划问题

整数规划

非线性规划

多目标规划

动态规划

随机规划

优选法

排队论

存 贮 论

对 策 论

决　策　论

图与网络

统　筹　法

模　型　论

军事运筹学

系　统　理　论

一般系统理论

动　态　系　统

突变与分岔

耗散结构理论与协同学

混沌与分形

决策与博弈系统

控制论系统

离散事件动态系统

不确定性系统

并行计算系统

系统工程

控 制 理 论

经典控制理论

线性系统

奇异摄动控制系统

广义线性系统及其他

随机控制系统

分布参数系统理论

最优控制理论

鲁棒控制

绝对稳定性

非线性系统

变结构控制

大规模系统

微 分 对 策

离散事件动态系统

计算机辅助设计

通信与信息理论

通信总论

数字通信

信　　道

保密通信

无线通信

电　　视

有线通信

光纤通信

信息理论

数字信号处理

网络分析

画法几何与工程图学

投影法与投影变换

曲线曲面

立体及其截断与相贯

展　开　图

轴测投影图

透视投影法

阴 影 法

工 程 制 图

计算机绘图

图 算 学

计 算 机 科 学

计算机科学的数学基础

计算机体系结构

计算机软件

人 工 智 能

机器证明

数理语言学

代数语言学

统计语言学

计算语言学

力　学

力学总论

理论力学

流体力学

弹性塑性理论

材料力学

结构力学

结构动力学

断裂力学

天 文 学

天体测量学

天体力学

天体物理学

射电天文学

空间天文学

测 绘 学

普通测量学

大地测量学

实用天文学

摄影测量与遥感学

工程测量学

海洋测绘学

地图制图学

测量误差及平差

测量器具

数　　学

数学(mathematics)　数学一词来自希腊文 μαθηματιχη',其字根 μα'θημα 意义为知识、科学,它非常恰当地反映这个领域的广泛性与普遍性.从历史上看,数学常常用其某个侧面来表示:中国古代用算学来强调其计算技术方面,而西方多用几何学一词代表数学,以显示欧几里得(Euclid)的《几何原本》传统,而实际上,其中也包括数论和量论的内容.随着时间的流逝,数学的内容不断地扩大,在17—18世纪直至19世纪,被包括在数学领域内的许多学科和分支已经独立出去,而在各学科的边界又不断创造和衍生出一系列新的学科,这些新学科现在已融合而成面向21世纪的庞大的数学科学领域,它是一个具有内在统一性的科学技术群.以下从四个方面进行论述:

1.数学的对象和特点.数学中最原始的对象是数与形.自然数已经是相当抽象的概念,它不仅要从一个苹果、一间房子、一堆沙土中抽象出数1来,而且还要由数1得出更一般数的概念.有了自然数的概念还会遇到基数和序数的矛盾.至于记数法和位值制都是中国对人类文明的伟大创造,这种伟大创造绝不仅仅是对自然界的认识和对哲学思辨的产物,它真正体现数学的成就.数学另一个原始对象是形,它更为直观,甚至长期以来人们也把它当成自然科学的对象,尽管柏拉图(Plato)早就说过,三角形属于理念的世界.当然现在数学的空间远远超出现实的空间,数学中的"形"也不限于人们感官能摸得着、看得见的东西,它是更抽象的概念,如高维空间、无穷集合、群、拓扑等是任何其他学科都不研究的对象.数学作为一门模式科学,应该归入更广泛的符号和形式科学类.这一类似乎应该介于哲学类与具体科学,即自然科学与社会科学之间.它的姊妹学科包括一般符号学、语言学、逻辑学、方法学以及还未成型的一般系统学.有意思的是,有些数学家也认为"数学是一种语言"、"数学可还原为逻辑"、"数学是一种普遍方法"等,这些说法尽管有些偏颇,但毕竟触到数学与自然科学的本质差别以及数学与符号科学的亲缘关系.数学的本质特征是:

1) 数学是一种普遍语言.这种观点可以追溯到莱布尼茨(Leibniz,G.W.),他首先提出科学与哲学的两大目标,其中第一个就是找出一种普遍文字,首先是一种符号及变元表示的符号语言.正如吉布斯(Gibbs,J.W.)所说的"数学是一种语言".吉布斯不仅是19世纪最伟大的统计热力学大师之一,而且也是向量分析的开创者及传播者.在19世纪90年代,英国著名的杂志《自然》上掀起的一场大辩论中,向量最终取代四元数而成为物理学普遍使用的概念.19世纪和20世纪之交,向量分析成为数学、物理学的有效工具,更确切地说,成为描述各种现象的语言.数学概念的产生及其符号化反映了数学的进步,算术运算的符号化及向符号代数过渡,几何学的代数化,微分、积分运算的符号化,函数的符号及行列式、矩阵、向量、张量等概念的符号化,复数的表示,算子演算以及符号逻辑等都是数学的重要进展.在这个意义下数学对象是一个符号集合.单纯的符号集合,正如裸的集合一样,没有结构,没有什么可说的,没有什么意思,而只有它具有形式结构(语法学),有一定的解释(模型——语义学),有一定变换、生成、操作、运用方式(语用学),它才能变得丰富多彩起来.数学作为一种普遍语言有自己的特点,比起纯逻辑语言来有内涵的丰富性,而比起通常语言来有外延的确定性.数学不仅是一种语言,它还是一种精密语言.正因为如此,它常被称为精密科学.数学之所以精密,不单是因为其数量表示,还在于它越来越深入那些以前所无法表示的或非实在的概念,如瞬时速度、加速度、位势、熵、谱等.对于许多直观概念,也只有在数学上才能得到很好处理,如连续性、对称性、随机性乃至信息控制、策略、对策、决策等.数学还明确了一些对立的范畴,如有穷与无穷、连续与离散、局部与整体、确定与偶然等.还有重要的元概念:如结构、构造、存在、模型、等价等.这些语言越来越深入到科学乃至日常生活之中,使论述确切及精密.许多常用概念也只有在数学上得到澄清才算有深刻的认识.

2) 数学是一种普遍方法.从古到今,对许多问题求出解答的过程中,人们或多或少产生一种方法的意识.而这种方法概念,又以数学中的算法概念为摹本.在这个意义下,数学充分显示其作为操作技术的特性.虽然精确的算法概念一直到20世纪30年代才有确切定义,但模糊的概念很早就有,而且也是数学追寻的主要目标之一.中国的数值运算及方程求解、欧几里得辗转相除法以及印度、阿拉伯的一些算法,都使人认识到算法是一种有限的指令,可以机械地运行,从而对一类问题得出确定的解答.许多几何作图问题以及求积问题也要求发现广义的"算法"来求解问题.笛卡儿(Descartes,R.)把算法推向普遍方法论的高度,他十分明确地考虑造出普遍方法

来解决科学问题，特别是数学问题，对此他称为普遍数学. 正是这种对方法的普遍考虑使他发明代数方法研究几何问题，从而创立解析几何学. 波利亚(Polya,G.)把笛卡儿的方案总结如下：

任意问题

$\xrightarrow{①}$数学问题

$\xrightarrow{②}$代数问题

$\xrightarrow{③}$解方程组 $\begin{cases} P_1(X_1,X_2,\cdots,X_n)=0, \\ P_n(X_1,X_2,\cdots,X_n)=0 \end{cases}$

$\xrightarrow{④}$解方程 $P(x)=0$,

其中第③,④步无非是数学，第②步并没有一般方法，第①步则困难更大. 莱布尼茨也把找出能求解任何数学问题的普遍算法列为他第二大目标. 笛卡儿在几何方面实现这个目标，而莱布尼茨则在逻辑上制定一个方案. 数学作为一种普遍方法，总是不断跨越已有的领域，深入到未知的世界中去，并且不断创造新的数学对象. 17—18 世纪，无穷及无穷小进入数学，把代数运算规则向无穷领域推广，这就导致数学分析的形式，并且构成方法上的大飞跃. 19 世纪由实分析向复分析过渡，再次加大分析方法的威力.

3) 数学是一种普遍思想原则. 数学发展过程中由于不断一般化、不断抽象化造成自己的普遍思想原则. 对称性原则、不变性原则、守恒律三者统一是 1918 年诺特(Noether,(A.)E.)首先提出的，而群论方法在量子力学、原子-分子结构、核结构、基本粒子理论中的适用性也是外尔(Weyl,C. H. H.)首先得出的. 群论方法至今仍广泛应用在科学的各方面，而不仅仅是上述领域及固体物理. 数学中另一个重要原则是极值原则或变分原理，最早除了哲学的思辨之外，首先提出的是费马(Fermat,P. de)，这些原则也是其后许多物理理论的基础，如哈密顿原理.

4) 数学是一种理性思维框架. 20 世纪之前，科学的支柱主要是理论和实验(包括观察和测量)，数学和计算包括在理论当中. 实际上，17 世纪的科学革命推动近代科学的产生，完全依赖于理性与经验的结合. 它们的哲学根源是笛卡儿的唯理主义与培根(Bacon,R.)的经验主义，他们也是近代哲学之父. 确切地讲，笛卡儿把认识论置于本体论之上，把哲学从神学的奴仆地位解放出来，成功地实行思想的解放，直接推动近代科学的产生，其中，理论概念、数学工具与观察实验结合在一起是牛顿(Newton,I.)科学革命的催生婆. 牛顿成就其伟业不仅在于他提出正确的理论概念(特别是力)，而且在于他提供了数学工具(微积分)及分析框架，尽管当时还是用欧几里得的几何系统. 其后科学的重大进步，理论概念及数学表述和计算的结合仍是不可缺少的一环. 第二次世界大战以后，计算机的发展与计算技术的进步使得科学计算与理论和实验鼎足而三，并列为科学的三大支柱. 近年来，由于数学的发展，从数学出发的理论越来越多地成为科学理论形成的始作俑者. 这特别表现在 1974 年，纤维丛理论成为规范场理论的标准表述. 其后越来越多的前沿数学领域进入物理学及其他科学领域，形成新兴理论框架，实际上数学已成为科学发展第四根支柱. 对于生命科学、心理科学、社会科学，这种现象早已不是新事物. 数理经济学以及对策论是这方面的最典型的例子.

2. 数学的分科及其主要问题. 数学不是一般意义上的自然科学或社会科学，它的对象及研究目标不像这些学科那样明确和集中. 从古到今，数学中所包含的对象、学科及分支变化多端. 中世纪除了算术及几何学之外，天文学及乐理也是数学的分支. 到 17 世纪，木工、石工、建筑、火器、占星术等都是数学的内容. 从那时起，静力学、动力学、光学、地图绘制法等仍然被看成大数学的一部分，尽管它们早已成为独立学科. 数学内容的庞杂也可以说是数学的一大特征了. 除此之外，许多基础的数学学科，它们的内容也有很大的改变，甚至于面目全非了. 经典代数学主要研究代数方程的求解，而经过几次变化，现代代数学主要研究代数结构. 这样一来，数学的统一尽管多次被提起，但是总难以概括全部数学. 因此，时至今日，数学仍然是具有多样性对象也具有多种目标的学科，尽管它们之间有着千丝万缕的联系. 人们把数学归结为相互关联的六大范畴，其中前三个可以说是数学的技术方面，后三个可以说是数学的理论方面.

1) 操作技术. 大部分最早的数学问题属于解决“如何”的技术问题. 最初的问题包括计数、计算、测量、作图等方面，后来逐步形成特殊的及一般的数学问题. 在解决这些问题的过程中，形成了算法以及操作步骤的概念. 在计算过程中，形成了算术，特别是解数值代数方程的算法. 到近代，这推动符号代数学、求解代数方程的技术以及把这些技术推广到无穷算法以及代数综合方法的代数方法，从而形成无穷小演算及解析几何学. 其后各个数学分支也提出相应的算法问题，例如拓扑学中计算同调群、同伦群等. 从这个意义上讲，数学在本来意义上是一种计算技术，或更广一点讲是操作技术. 而研究这种技术的目标就是发明算法或解题的步骤，以求得问题的解决. 应该说，这是一种富有创造性的研究工作. 以计算为例，就是由精确计算到近似解析计算到数值计算到计算机软件，它一直是数学研究的重要内容. 除计算之外，还有测量、绘图、统计、运筹等操作，以及相应的和衍生的各种问题. 例如古典几何有许多几何作图问题，特别是用圆规、直尺的几何三大问题，

以及更一般的作图方法. 为了解决这些问题,还要发明许多技术,如各种投影技术,它们至少在过去都属于大数学范围之内. 在数学分析的范围内,级数求和、渐近展开、积分变换等都是高级的计算技术.

2) 技术理论. 对数学操作的对象,应该有些认识,其中包括表示问题、操作规则与规律问题、可计算性问题、无穷级数收敛与发散问题、收敛速度问题、方程可解性问题、逼近的程度及可能性问题、作图的可能性问题,特别是方法的评价问题等. 这样就形成与操作技术有关但又高一层次的学科,如数值分析、误差理论、函数逼近论、丢番图逼近理论、可解性及稳定性理论等.

3) 操作对象理论. 它的目标不是指向对象本身,而是指向技术,指向求解的方法. 例如丢番图方程论、代数方程论、代数方程组理论、常微分方程论、偏微分方程论、积分方程论等. 它们涉及的不是单个方程,而是一类的对象,因此首要问题是分类问题. 然后再对每一类研究解的数目、解的性质、根与系数的关系、某类解的存在性问题等,微分方程的定性理论也属于这一范畴.

以上三大范畴是解决"如何"的技术问题,而以下三大范畴才是解决"什么"的理论问题.

4) 对象理论. 理论是对确切定义的对象的性质、关系、刻画、分类等的研究. 典型的数学理论有数论、函数论、算子论以及各种几何对象理论、过程理论等. 以数论为例,重要的分支按对象分有整数论、代数数论、超越数论等. 按方法分有初等数论、解析数论、概率数论等. 按问题的性质分有型的算术理论、几何数论等,也包括数论中的丢番图方程理论. 函数论与算子论一开始也是表示问题,特别是无穷级数及无穷乘积表示,然后是积分表示等,其次涉及值分布等可以说是计数问题,另外还有刻画及分类问题. 几何图形有许多性质与关系方面的问题,如度量性质以及相交、属于等关系,也有刻画及分类问题.

5) 结构理论. 结构理论与对象理论之间并没有一条不可逾越的鸿沟,这样划分是因为结构理论必须建立在集合的基础之上. 按照布尔巴基学派的观点,原始的结构可以划为三大类,研究它们各自的结构就形成结构数学的主要分支:① 代数结构:主要是群、环、域、模,它们分别构成群论、环论、域论、模论;② 序结构:主要是格,它们构成格论;③ 拓扑结构:主要是拓扑空间,它们构成一般拓扑学的研究对象. 这些抽象的研究对象有两个来源:一是从过去研究的具体对象抽象化,特别是公理化而成,如群、域以及拓扑空间这些抽象结构衍生出来. 新结构的产生有如下的几种途径:① 增减公理;② 复合结构;③ 多重结构;④ 混合结构. 研究这些抽象对象的目标是搞清楚它们的结构并加以分类. 所谓结构,就是元素(或它们的集合)和元素之间的关系. 结构数学的主要问题大致可分为互相关联的四类问题:① 刻画问题;② 分类问题;③ 结构问题;④ 实现问题.

6) 元理论. 元数学理论是对数学本身进行反思的产物,长期属于哲学的范畴. 它讨论数学概念、数学理论的合理性以及数学方法的合法性问题. 19 世纪末之前,对于数是什么以及非欧几何问题,特别是数学分析的严格性的争论均属于这个范畴. 19 世纪末,集合论的建立,现代公理方法的提出,符号逻辑的形成,以及关于数学基础问题的论战,最终导致作为一门数学分支的数理逻辑的形成. 由于哥德尔(Gödel,K.)的工作,数理逻辑成为包括模型论、公理集合论、递归论和证明论(原始的和狭义的元数学)四大分支的数学领域,其后分别形成构造性数学和计算复杂性理论等新兴学科. 除了以集合论为基础的现代数学之外,范畴论也成为一门元理论,在数学中有着有效的应用.

3. 数学的发展和演化. 数学的内容及范围随时间不同而不同,因此有言:"数学无非就是它的历史."数学史大致可如下分期:前史时期;古代及中世纪时期(从公元前 4 世纪到 16 世纪末);近代前期(17—18 世纪);近代后期(19 世纪);现代时期(20 世纪). 每一个时期的特点简要分析如下:

1) 前史时期. 前史时期的数学主要是民族数学或文化数学,在各种文化的发展过程中,各民族都或多或少掌握一些简单的数学技术,包括计数、计算、测量、土木建筑、绘图等,基本上属于实用技术. 这些知识是零散的,而且反映出较大的文化差异. 另外,也出现了神秘的占星术、数秘术、占卜术等,其中有个别涉及数学的内容,如二进制等. 各种建筑上的对称图案以及正多面体的列举包含群的观念的萌芽.

2) 古代及中世纪时期. 数学经过长时期的发展之后,正式成为一门学科,其主要标志是:① 建立数的表示及计算方法;② 对于一些问题有较系统的方法. 这使数学技术部分初步形成. 而欧几里得《几何原本》的问世,则使理论数学有了一个原型. 各国数学发展状况有所不同,古代数学的主要领域是算术与几何,希腊具有初等的数论及量论以及一些基本的几何问题及数论问题. 这些问题对以后的数学发展有很大的影响,但不一定很重要,比较重要的数学是计算,特别是解方程. 中国、印度、阿拉伯的数学,偏重于计算及实际问题的解决.

3) 近代前期. 近代数学诞生的标志是符号化的普遍算法的建立以及无穷进入数学. 它一下子使建立在几何及算术上的算法登上了一个新台阶,不仅使它的应用范围大大扩大,成为发展科学技术的有力工具,而且也向理论提出一系列问题. 这就导致

19 世纪操作理论、操作对象理论、对象理论的产生，出现数学多样化和理论化的时代. 17 世纪符号代数、解析几何学及微积分的建立，虽然大大扩大了数学技术库，但是并没有改变数学主要是一门实用的计算及操作技术的状态. 数学作为一门计算技术进步惊人，特别是微积分的完成，解决了许多天文、力学及物理学的问题. 微积分本身只不过是一种更有效的演算方法，即所谓无穷小演算. 接着是常微分方程及数学物理方程的出现，以及变分法的诞生，使数学工具更为有效.

4）近代后期. 19 世纪数学是近代数学的成熟时期，也是数学真正作为自为的理论科学产生的时期，但是伴随操作理论（如最小二乘法及误差理论、级数求和理论、函数逼近理论及丢番图逼近理论）、操作对象理论（如代数方程理论、常微分方程理论），数学技术本身也大有提高，特别是傅里叶展开、积分变换，尤其是复分析的建立. 19 世纪的数学可以说是数学对象化与多样化时期，一方面把数学由主要是操作技术转变为理论的时期，另一方面也为 20 世纪现代数学奠定了基础. 这样，数学对象理论真正形成，数学成为一种自为的科学而不再仅是自然科学或技术的语言和工具了.

5）现代时期. 20 世纪的数学是从 19 世纪数学多样性时期趋于统一的时期，其统一的基础是集合论. 一方面集合论之上产生了结构数学的庞大领域，另一方面集合论的基础问题产生了元数学. 数学新对象的形成，产生结构的多样性，导致理论的多样性，并且 19 世纪末以前的四大范畴的数学仍有新的发展，加上新的应用数学、计算数学等领域，数学日趋专门化、多样化. 但意想不到的是，从 20 世纪 70 年代起，各个领域之间新关系不断发现，新一轮的统一性正在形成之中. 当代数学前沿的大多数学科是 20 世纪上半叶形成的，其中主要是抽象代数学（包括群论、环及代数理论、域论、格论、整体李群理论、代数群论、同调代数以及各种衍生结构）、一般拓扑学、测度和积分理论、泛函分析（包括线性拓扑空间理论、算子代数理论等）、组合及代数拓扑学、整体微分几何、多复变函数论、动力系统理论、随机过程理论等. 对于 19 世纪开创的新领域——代数数论、代数几何学、黎曼几何学和局部李群理论，也在结构数学的框架中获得重大突破，成为当代数学的前沿.

20 世纪后期形成的一些领域，如微分拓扑学、大范围分析、K 理论、非交换几何等，也可在其中看到其萌芽. 除了纯粹数学领域的扩大与深化之外，20 世纪的应用数学和计算数学的面貌也发生了根本的改变. 一方面数学应用的范围已从 20 世纪之前的经典力学、天文学与测地学以及数学物理等领域扩展到几乎所有自然科学、工程技术、社会科学、人文科学的分支，并越来越起着举足轻重的作用；另一方面，一批新的应用数学领域产生出来，成为具有相对独立的分支，构成大数学科学的组成部分. 它们一方面与实际问题有密切的关系，另一方面它们也形成独立的数学研究方向，其中最典型的是 19 世纪末 20 世纪初形成的数理统计，它们同应用概率一起在近半个世纪已经成为与经典数学平起平坐的学科领域. 另外一个数学领域——组合数学几乎与数学的历史一样悠久，但只是近半个多世纪才逐步成熟及独立. 第二次世界大战之后，一些新的应用数学领域独立出来，特别是运筹学诸分支，后来纳入管理科学的学科群中. 与工程技术密切相关的系统科学、控制理论与自动化科学、信息科学也得到空前的发展.

20 世纪科学技术史中头等重要的事件是电子计算机的诞生. 它对整个社会的冲击是怎么估计也不过分的. 从计算机的设计制造到大规模应用，处处离不开数学，同时也开辟了新的数学领域. 它们可以归纳成两大部分：一是计算机科学，它指未来计算机的发展；一是计算数学，它指计算机在科学计算和工程技术中的大规模计算. 计算机的不断普及和改进对数学也造成不可忽视的影响. 它给数学家提出一系列算法问题，并形成一套行之有效的算法，如单纯形方法及其种种改进，有限元方法及其衍生算法等，对算法的分析，如收敛速度、误差传播及稳定性等问题形成数值分析分支. 近年来，计算机由数值运算过渡到符号运算，形成计算机代数重要分支，特别是吴文俊的机械化数学纲领在机器证明方面是一大突破.

4. 数学的社会功能. 数学是最古老的科学部门，它的诞生和发展反映人类文明的进步. 数学从一开始就与社会实践活动密切相关，从计数、土地丈量、器物制造、产品分配，一直到商业贸易、宗教活动等都向数学提出问题，并要求逐步解决和方法逐渐进步，最后形成相对定型的数学方法和学科. 从此，各种社会活动与数学的应用密不可分. 随着社会的进步，特别是近代科学技术的进步和新兴产业日新月异，数学也越来越成为科学技术发展的基础. 从 17 世纪到 19 世纪，数学与力学、天文学、物理学、大地测量学、航海术就密不可分，互相促进地平行发展着. 对于机械工程、建筑工程设计、电机工程等技术领域的发展，数学也起着决定性的作用. 20 世纪数学科学的巨大发展，比以往任何时代都更加令人信服地确立了数学作为整个科学技术的基础地位. 数学物理、数学化学、生物数学、数理经济学、数理地质学、数理语言学、数值天气预报、数学考古等一系列边缘学科的出现，表明数学的应用已突破传统的范围而向人类一切知识领域渗透. 随着科学数值化趋势的增长，数学在提高全民素质，培养适应现代化需

要的各级人才方面还具有特殊的教育功能. 数学科学,已成为推进人类文明的不可缺少的重要因素,数学正越来越直接地为人类生活与物质生产做出更大的贡献. 数学应用具有以下特点:

1) 纯粹数学几乎所有的分支都获得应用. 在20世纪60年代,像拓扑学这样的抽象数学分支离实际应用似乎还很遥远,而今拓扑学(特别是扭结理论)已成为生物学中了解 *DNA* 结构的有效工具. 在物理学中,拓扑不变量正在成为物理的量,正如一些群的不变量是物理的量一样. 数论也曾被认为是最纯粹、最缺乏应用的数学分支,但如今数论方法在计算机科学、密码技术、卫星信号传输、p 进量子场论等许多方面发挥着重要的有时甚至是关键的作用,并通过与数值分析相结合开辟着更广的应用途径. 事实上,仅就在理论物理中的应用而言,涉及的数学除了经典的分支与方法(如数学物理方程、傅氏分析、无穷维空间论、群论、概率统计等),还包括了微分拓扑、微分几何、大范围分析、代数几何、李群与李代数、算子代数、代数数论、非交换数学、非线性数学、计算数学等,几乎覆盖了核心数学的整个领域.

2) 几乎所有的科技领域都在应用数学,并越来越多地应用更高深的数学. 数学在力学、物理学中的应用是经受了历史考验的,而当今数学的应用则早已突破这一传统的范围,正在向包括从粒子物理到生命科学,从航空技术到地质勘探在内的一切科技领域进军. 除了自然科学,经济学及过去认为不适用数学的社会学、历史学等社会科学领域,数学方法也都在崭露头角. 随机分析应用于金融决策而引起的经济学理论的进展,提供了特别令人鼓舞的例证. 与以往时代不同的是,数学在向外渗透过程中越来越多地与其他领域相结合而形成交叉学科. 与数学有关的词大量出现在各门学科之前后,如"数学的"、"数理的"、"计量的"、"统计的"、"计算的"以及"……数学"、"……统计学"等. 学科成熟的社会标志是学会、协会的建立,期刊与连续出版物的问世,以及课程的设置,专业会议的召开等. 例如,《数学化学杂志》于20世纪80年代创刊,《数理经济学杂志》于20世纪70年代创刊,生物数学的期刊出现更早. 次一级的学科如"数学分类学"的著作早在20世纪80年代就问世了. 值得注意的是纯粹数学中的一些前沿与其他科学的许多前沿领域的快速结合,这反映了科技领域中数学渗透的空前深度. 可以这样说,没有这些前沿数学就没有当代理论物理学的一些前沿领域,如超弦理论、超引力理论等. 事实上,仅仅像弦理论这样的物理学热门分支所用到的数学,就涉及微分拓扑学、代数几何学、微分几何、群论与无穷维代数、复分析与黎曼曲面的模理论等. 凝聚态物理中分类晶体结构中的"缺陷"以及液晶理论,都用到某些齐性空间中同伦群的计算,而这即使对代数拓扑学家来说也是极难的问题. 数理经济学中一般均衡理论的建立、发展,也用到了微分拓扑学的基本定理与彻底的公理化方法. 经济学家德布鲁因(de,Bruijn,N. G.)这方面的工作获得了诺贝尔奖.

3) 数学在生产技术中的应用变得日趋直接. 以往数学工具直接用于生产技术的例子虽有发生,但数学与生产技术的关系基本上是间接的. 常常是先应用于其他科学,再由这些科学提供技术进步的基础. 近半个世纪来,数学科学与生产技术的相互作用方式正在悄悄地改变,数学提供的工具直接影响和推动技术进步的频率正在加大,并在许多情况下产生巨大的经济效益. 例如,以计算流体力学为基础的数值模拟已成为飞行器设计的有效工具,类似的数值模拟方法正被应用于许多技术部门以替代耗资巨大的试验;以调和分析为基础发展起来的小波分析直接应用于通信与石油勘探等广泛的技术领域,这在20年前是不能想象的;现代医学扫描技术(*CT* 扫描、核磁共振成像等)主要也是建立在拉东积分理论的基础之上,这方面的例子举不胜举. 此外,现代大规模生产的管理决策、产品质量控制也密切依赖于数学中的线性规划算法(单纯形法与新兴的内点法)及统计方法. 近年来,以数学建模为核心的工业数学成为一个蓬勃发展的应用数学领域也绝不是偶然的,产业部门的工程技术人员与数学工作者携手合作,解决影响甚至决定生产过程的形形色色的数学问题,反之,许多挑战性问题也刺激纯数学的发展.

4) 数学在学科发展中的份额及力度越来越大. 一些著名的数学家认为,数学是一种关键的、普遍适用的、赋予人以能力的技术. 从某种意义上来讲,"高技术本质上是一种数学技术". 对此,一般人还只在科学计算的层面来理解. 而实际上,数学方法是不同于理论方法及计算方法的第四个普遍适用的方法和技术. 这种情况从20世纪70年代以来已初露端倪,在21世纪将成为科学研究的重要组成部分,而且也许是最富创造性的部分,这特别表现在形成概念及理论框架方面. 实际上,当前的动力系统的研究(分叉、吸引子)、孤立子、混沌等已成为许多领域的通用语言及工具,而更艰深的数学将在未来更为普及.

执　笔　胡作玄
审　阅　吴文俊　程民德　徐利治

应用数学(applied mathematics)　应用性较强的诸数学学科或分支的统称. 也有一种理解为,泛指一切数学理论和方法中应用性较强的部分. 这里的"应用"是指用于解决数学以外的生产实践、科学技术以及各种社会活动中的理论或实际问题. 从人类

社会发展史上看，数学本来起源于实际应用的需要，应用一直是数学的发展动力之一，一种数学理论和一门数学学科的生命力的强弱，很大程度上依赖于它有无应用.因此，从这个意义上讲，整个数学都是有用的.事实上，古老的算术与平面几何等不但起源于实际应用，在现代社会活动中也是不可一时或缺的.许多昔日认为毫无实际意义的数学理论，如数论、拓扑、凸分析以至非标准分析，都在生产和科技发展中找到它们的应用.因此不能简单地以“有用”或“无用”来区分应用数学的学科.像微积分在天文、物理及许多科技领域都有广泛的应用，理论力学就完全是应用微积分解决经典力学问题的学科.但从整个学科看，应用只是它的一部分，以微积分为其主体的数学分析是分析数学的基础，所以它还是应该属于纯数学.

在古代，数学本无“应用”和“纯粹”之分，“应用数学”一词大约出现于17世纪前后，高等数学方才萌芽，数学的抽象层次提高，理论和应用开始明显分离的时期，从此人们用它专指数学的某些部分，以至某些独立的数学学科.所以，应用数学学科一般都是在近代或现代发展起来的，其特点是，它们的内容都是围绕着解决某一类实际问题，因此首先要有从实际问题中抽象出相应的数学模型，使其具有应用数学理论和方法的基础；其次，所提出的解决问题的方法是可行的，而不是只从理论上证明问题解的存在.例如，牛顿(Newton，I)提出地球绕日运动的模型，是用微积分方法解决天文学问题的开端；阿罗(Arrow，K，J.)与曾获诺贝尔经济奖的德布罗(Debreu，G.)给出的一般经济均衡模型，使得经济学问题中可以使用的数学方法不止微分法，而涉及更多的数学领域，这才形成数理经济学学科.当然，应用数学学科也并非从各数学分支抽取一些有用的方法简单堆砌而成，而有其自身的系统理论，模型和方法都有其理论依据，否则也不成其为独立的学科.

应用数学一方面应用于各类实际问题，另一方面，它也是纯数学的应用，纯数学是它的理论基础.像数理统计是从已知的关于有限的样本的信息出发，去研究无限整体的统计规律，它给出由样本估计整体的统计量的统计方法，同时对这些方法的可靠性、优劣标准等有系统的论证.这些方法和论证主要以概率论作为理论基础和根据，又如微分和凸分析主要研究极值问题的思想和准则，而运筹学则着重研究最优化问题的种种解法.从研究对象上看，应用数学也不都是直接研究实际问题，它的对象主要是数学模型，用的还是抽象的数学方法，提供的是解决问题的思维准则.这些都与纯数学相近，而与工程技术或技术学科不同.然而，纯数学所追求的，有时可以只是理论上的完善和方法上的优美，一般并不是针对某一类实际问题，例如，数论中对哥德巴赫问题的研究就是如此.应用数学追求的目标一般则是解决模型所提出的问题，而模型是有实际意义的.因此应用数学与纯数学在其研究的目的性方面大不相同，这是二者的最大区别.

应用数学所包含的学科难以胜数，其领域随着时代的发展正日渐扩大.有人甚至将理论力学、天文学和测绘学等与数学关系密切的学科也归入其中.各种各样的计算方法，显然也与数学的应用有密不可分的关系.特别是近年来，随着电子计算机技术的快速发展，与之相应的各种计算方法变得名目繁多，有了长足的进步，由于其专用于计算的共同特点，而不能继续附属于各数学学科，而属于新兴的计算数学.在应用数学中，较早形成学科的，有基于分析和微分方程方法研究现代物理问题的数学物理，和运用概率统计方法探讨随机现象的数理统计、生物数学等.其后的发展，这些学科内容更加丰富，还生成了许许多多的新学科，诸如综合运用概率统计及其他数学方法，总称运筹学的数学规划，排队论等，20世纪中期以后发展起来的控制论、信息论、系统论、模糊数学、突变理论、小波分析、混沌理论、分形几何以及组合学方面的学科也都整体或部分地属于应用数学.近年来，数学的应用更深入到社会和人文科学领域，数理经济学、金融数学等名称冠以“数学”或“数理”的新学科如雨后春笋般涌现，无不显示出应用数学发展的旺盛活力.

目前人类已进入科技和生产高速发展的时期，它一方面使应用数学大有用武之地，同时又对数学，主要是应用数学提出了大量的需求.人类无论在对揭示生命起源等微观世界的研究，或是在对展现宇宙奥秘的宏观世界的探索中，都需要数学的帮助.因此应用数学的未来发展前景是十分美好的.

撰　稿　常心怡　　**审　阅**　萧树铁

运　筹　学

运筹学(operations research)　一门应用十分广阔的数学边缘学科.它研究一给定的运行系统如何在某种意义下达到最优运行的问题.运筹学研究问题的主要方式是:建立相应的数学模型,然后寻求该模型的数值解答.于是,产生了各种运筹数学理论,形成若干运筹学的数学分支.运筹学研究的问题大多以最优化形式出现.它的理论和方法具有十分广阔的应用范畴,并且是现代决策科学、管理科学和系统科学的重要基础.

运筹学起源于第二次世界大战初期.当时的英国(随即是美国)军事部门迫切需要研究如何将非常有限的物资(人力和物力)分配和使用到各种军事活动(运行)中,以达到最好的作战效果.1935年,为了对付德国空中力量越来越严重的威胁,英国在波德塞(Bawdsey)成立了关于作战控制技术的研究机构.1938年,波德塞科学小组负责人罗(Rowe,A. P.)把他们从事的工作称为运筹学(operational research[英],operations research[美],直译为作战研究).由此,人们把波德塞作为运筹学的诞生地,把1935—1938年作为它产生的酝酿时期.关于运筹学的一些先驱者的工作,则可以追溯到1909年丹麦电气工程师埃尔朗(Erlang,A. K.)有关排队论的早期论文,以及1916年英国工程师兰彻斯特(Lanchester,F. W.)的战争模型等.但由于这些工作是孤立的,未形成系统,因此人们认定,作为一门学科的起源,运筹学是在第二次世界大战中诞生的.

在战争结束时,主要来自英国、美国和加拿大的运筹学工作者人数已超过七百,其中一部分人力图将他们在战争中取得的经验和知识转到民用生产中去.于是,在英国先出现了一个"运筹学俱乐部",1950年出版了第一份运筹学杂志,并于1953年成立了英国运筹学学会.在美国,则于1952年即成立了美国运筹学学会.此后,于1959年成立了国际运筹学联盟(International Federation of Operational Research Societies,简称IFORS).该联盟现有会员国45名.在中国,1956年中国科学院即组织开展运筹学研究.1964年,借鉴《汉高祖本记史记卷》中"夫运筹帷幄之中,决胜于千里之外"一语,把"operations research"译名为"运筹学".中国运筹学学会成立于1980年,1982年加入IFORS,并于该年创办了《运筹学杂志》,1997年改为《运筹学学报》.

20世纪50年代以来,由于大型企业的兴起,科技水平的提高,产品更新的加速和市场竞争的激烈,促使在经济和管理领域有着广泛应用的运筹学得到快速地发展.现在,全世界(尤其在美国)的一些大型企业多设有自己的运筹学研究组,在一些较知名的大学里,一般地,皆开设运筹学课程,不少大学还建有运筹学专业或有关的系和研究所.

大型的实际问题牵涉面一般都很广,因素也很多;用数学式子描述,有的问题中出现的变量数甚至要达到上百万.对此,传统的数学手段已无能为力,必须发展新的数学方法.1939年,苏联数学家坎托罗维奇(Канторович,Л. В.)在《生产组织与计划中的数学方法》报告中,提出了求解大型特殊极值问题的"解乘数法".1947年,美国运筹学家丹齐克(Dantzig,G. B.)则提出了求解线性规划问题的著名的"单纯形法".他们的工作开创了用数学方法解决大型实际问题的先河.由于生产实践中的大量问题皆可归结成线性规划模型,而单纯形法对于当时出现的线性规划问题的规模行之有效,这对早期运筹学的发展产生了很大影响,并起了重要的推动作用.随着运筹学应用领域的扩大,从运筹学诞生以来的50余年里,运筹学已经发展成包括有众多新的数学分支的庞大学科.

运筹学的一个主要和大的分支是数学规划.它研究在一些给定的条件(约束条件)之下,求所考察函数(目标函数)在某种意义下的极值(极小或极大)问题.运筹学中的数学规划问题与经典数学中的极值问题不同.因为数学规划问题来源于实际,其中的变量数一般都很大,并且目标函数常常十分复杂,甚至无法写出其具体的表达式.此外,研究数学规划所关心的是如何求得问题的数值解,而解答的数学表达式并不重要.因此,对于运筹学家来说,研究数学规划,重要的是要设计一些应用面广、容易编制计算机程序和收敛速度快的求解方法.当数学规划问题中的目标函数是变量的线性函数,并且约束条件可用变量的某些线性等式和(或)不等式来表示时,它即为上面提到的线性规划问题.对于求解线性规划问题,现今已有了改良的单纯形法、内点法等可以处理超大型模型的算法.若线性规划问题中的一些变量只限于取整数值,则称它为整数规划.求解整数规划问题的难度远比求解一般线性规划问题要大.若在数学规划问题的目标函数或(和)构成约束条件的函数中出现非线性函数,则此问题称为非线性规划问题.非线性函数的复杂性大大增加了解决问题的难度.1951年,美国运筹数学家库恩(Kuhn,H. W.)和塔克尔(Tucker,A. W.)建立了非线性规划问题最优解的必要条件,为该学科奠定了重要的理论基

础.自20世纪50年代以来,不少学者在非线性规划领域开展研究,取得了很大进展,至今它仍是运筹学研究中的一个热门分支.随着数学规划研究的推进,一些数学分支,如凸分析、非光滑分析,以及数值计算等也得到很大发展.

有一些运筹学问题,例如计算机线路的设计、最优工序的安排等,一般地,很难写成一个以算式表达的数学问题.有时即使写成了,由于引入的变量和约束条件过多,求解也很困难.对于这类问题,人们总是设法用组合的方法去求解,因此称之为组合最优化问题.由于对它没有统一的解法,一般地,是根据问题性质的相近,将其分成若干个以某一直观模型命名的问题.如排序问题、匹配问题、网络流问题等.组合最优化的一个特点是:多数问题属于所谓的NP完全问题,即对该问题基本上不存在一种算法,使得当所与的具体问题的变量和约束条件的数目两者之和甚大时,可以在容许时间(即所谓的多项式时间)之内给出所要的解.由于这类问题在生产实际中经常出现,不能予以忽视,于是出现了两类解决问题的途径:一类是所谓的直观算法,另一类是近似算法.随着组合最优化研究的进展,一些数学分支,如组合数学、拟阵和广义拟阵,以及图论等,也相应地得到新的发展.

生产中常出现一些问题,它们所涉及的一些输入信息随时间作微小变动时,目标函数的值可能随之引起大的变化.基于这种现象而产生的一个数学规划分支称为参数规划.其目的是研究当问题中出现参数时如何求解,以及解的性质和目标函数的值对于这些参数的依赖情况.

有些问题所考虑的目标不止一个,例如修建一座水库,要考虑的目标有航运、灌溉、防洪、防旱,以及安全等.有些目标甚至相互排斥.基于这类现象便出现了多目标规划.当然,这时最优值的涵义与单个目标的情况不同,它有着各种不同的定义.多目标规划也是运筹学的重要分支之一.

在以上所述的数学规划问题中,所涉及的信息和变量皆不含不确定因素.若其中出现随机变量,则称其为随机规划.它也是数学规划的分支之一.

在随机问题方面,运筹学中的重要分支有排队论、可靠性理论、决策分析等.在第二次世界大战中,由于飞机的大量出动,而机场的跑道很有限,因此对于飞机的维修、起飞时间的安排和跑道增设的估计等,皆须做出筹划.对此,一些科学工作者借用战前埃尔朗关于电话服务系统的理论进行处理,取得了效果.战后又再将其移到民用,逐渐扩大形成排队论学科.在现代社会,由于电子工业的发展,特别地,是一些复杂的高精产品所涉及的元件和线路非常多,某一元件的失灵即可影响全局,因而设备的可靠性自然成为重要的研究课题.此外,由于工业的发展,使得许多大型的、复杂的项目的实现成为可能.一个大项目的兴建(例如一个大地区的开发)涉及的面很广,在做出决定之前需要对整个项目进行周密的筹划.决策理论即是为决策者提供一种系统的、较全面的、合乎逻辑的方法,以帮助决策者做出好的抉择.

属于运筹学研究范围的还有对策问题、库存问题和搜索问题.对策论又称博弈论,它研究对于由一些相互具有竞争性的群体(或个体)所构成的体系中,每个群体在竞争的每一步如何在供选策略中选取一策略,以使在局终时能得到最好的结果.若每一策略的选取是按某种确定性规划做出的,则此种规则称为纯策略;若依某种概率来选取,则称之为混合策略.再根据局终胜负的定义之不同,以及局中人是否有合作关系等,使得对策论成为一门内容丰富的学科.对策论产生于二战后期.1944年,美国数学家冯·诺伊曼(von Neumann, J.)和莫根施特恩(Morgensten, O.)所著《对策论与经济行为》一书是其早期较系统的工作.库存论是研究物资储备控制策略的理论,即决定何时补充库存和补充多少.根据涉及因素的有无随机性、物资是否有失效期或变质、储存费用、缺货损失等,需要做出最优或较好的储备策略,库存论便形成了一门具相当规模的学科.搜索论所研究的是利用有限的资源和探测手段,如何设计一种方案,以最大可能(或最短时间)寻找到某种目标的问题.搜索论起源于二战,为了搜索大西洋中袭击盟军商船的德国潜艇而进行了研究.战后,搜索论在理论和方法上都有了相当规模的发展,并且在一些重要的项目中得到了成功的应用.

运筹学的应用如此广阔,要将一个实际问题形成一个或一组真正能反映问题本质的数学模型,然后予以解决,一般地,要求数学工作者要与问题有关的技术人员、管理人员等一起工作.有的问题由于过于复杂,无法形成一个可以利用现有手段解决的数学问题,有的甚至无法形成数学问题,这时就须借助其他的科学方法,如计算机模拟或实验手段来进行求解.

运筹学的另一特点是:必须有可能将问题的解答及时求出,否则事过境迁,于事无补.因此,运筹学的理论和应用的发展与计算机和计算技术的发展息息相关.

撰稿　越民义　　审阅　胡毓达

最　优　化

最优化(optimization)　应用数学的重要研究领域.它是研究在给定约束之下如何寻求某些因素

（的量），以使某一（或某些）指标达到最优的一些学科的总称. 由于运筹学中出现的问题大多即是最优化所研究的问题，因此运筹学的许多分支，如数学规划、组合最优化、排队论，以及决策论等也是最优化的组成部分. 此外，最优化还包括工程最优设计、最优控制（控制论与运筹学的交叉分支）等. 狭义地说，最优化即指数学规划，有时也专指非线性规划.

最优化问题（optimization problem）　最优化各学科研究的问题. 也是运筹学中大多数分支研究的问题. 在给定约束之下，从问题的许多可能解答中，寻求使某一（或某些）指标达到最优解答的问题. 用数学语言来描述，即是在一给定的集合上求某泛函的极值（极小化或极大化）问题. 对于极小化的情形，即为求

$$\min_{u\in(\subset)U} J(u)$$

的问题，其中 $u\in(\subset)U$ 表示变量或集合 u 要满足所给的约束（组成集合 U 的条件称为问题的约束条件），泛函 J 表示用以对解进行选择的指标（称为问题的目标函数）. 研究最优化问题有悠久的历史. 公元前 212—187 年，古希腊数学家阿基米德（Archimedes）就曾证明了已知周长，圆所包围的面积最大的等周问题. 中国古代城堡是方形的，这是因为给定周长时，正方形是包围面积最大的四边形. 在 20 世纪 50 年代以前，主要是研究求可导函数极值（微分法）和泛函极值（变分法）的古典最优化问题. 20 世纪 50 年代以来，产生了用计算机求解大型的最优化问题. 最优化问题可分为以下五类：

1. 无约束和约束最优化问题（没有约束条件限制的最优化问题称为无约束最优化问题，有约束条件的称为约束最优化问题）.

2. 确定性和随机性最优化问题（确定性最优化问题，即每个决策变量取值是确定的. 随机性最优化问题，即某些决策变量取值是不确定的，但知道决策变量取某值而服从一定的概率分布）.

3. 线性和非线性最优化问题（如果目标函数和所有约束条件中的函数都是决策变量的线性函数，这种最优化问题称为线性最优化问题. 如果目标函数或约束条件中至少有一个是决策变量的非线性函数，这种最优化问题称为非线性最优化问题）.

4. 静态和动态最优化问题（如果最优化问题的解不随时间而变，这种最优化问题称为静态最优化问题. 若问题的解随时间而变化，则称它为动态最优化问题）.

5. 单目标和多目标最优化问题（如果问题中只含有一个数值目标函数，这种最优化问题称为单目标最优化问题. 若问题中的目标函数多于一个，则称它为多目标最优化问题）.

确定性最优化问题（deterministic optimization problem）　见“最优化问题”.

随机性最优化问题（stochastic optimization problem）　见“最优化问题”.

线性最优化问题（linear optimization problem）　见“最优化问题”.

非线性最优化问题（nonlinear optimization problem）　见“最优化问题”.

静态最优化问题（static optimization problem）　见“最优化问题”.

动态最优化问题（dynamic optimization problem）　见“最优化问题”.

单目标最优化问题（single objective optimization problem）　见“最优化问题”.

多目标最优化问题（multiple objective optimization problem）　见“最优化问题”.

组合最优化问题（combinatorial optimization problem）　一类特殊而重要的最优化问题. 即在给定有限集合的所有具某些特性的子集簇中，寻找使某种指标达到最优的子集的问题. 依据问题的性质，它包括排序问题、匹配问题和网络流问题等.

网络最优化问题（network optimization problem）　一类特殊的组合最优化问题. 应用图论的理论，通过网络的拓扑结构及其性质，对网络进行研究，并用计算机算法寻求网络中的最短路、最大流等，这种最优化问题称为网络最优化问题.

最优化方法（optimization method）　求解最优化问题的方法. 也即从最优化问题的可能解答中，寻求使某一（或某些）指标（目标）达到最优的解答（最优解或满意解）的方法. 用最优化方法解决实际问题时，一般分三步进行：

1. 提出最优化问题，建立数学模型.

2. 分析模型，选择合适的求解方法.

3. 编制程序，用计算机求最优解，对算法的收敛性、通用性及误差等做出评价.

解析最优化方法（analytic optimization method）　亦称间接最优化方法. 求解最优化问题的一类方法. 它要求把所研究的对象用数学表达式描述出来，然后用数学解析方法求出最优解.

间接最优化方法（indirect optimization method）　即“解析最优化方法”.

直接最优化方法（direct optimization method）　亦称试验最优化方法. 求解最优化问题的一类方法. 对某些实际问题，直接通过有限的试验或模拟，并将结果加以比较，从而获得最优解或近似最优解. 有些难以用数学表达式描述的实际问题，常常采用这种方法.

试验最优化方法（experimental optimization

method） 即"直接最优化方法".

组合最优化方法(combinatorial optimization method） 求解组合最优化问题的方法. 一般地，对于不同类的组合最优化问题，对应着不同的求解方法. 判定一个组合最优化方法好坏的主要标准是运算次数. 用 n 表示某一组合最优化问题的规模，$p(n)$表示在对方法影响最坏的情况下所需的运算次数. 若 $p(n)$是 n 的多项式函数，则称该方法是多项式算法. 凡能用多项式算法求解的问题都称为 P 问题. 有一类问题称为 NP 完全问题，若这类组合最优化问题具有如下特点：

1. 它们都未找到多项式算法.

2. 如果对其中某一问题存在多项式算法，那么此类中的所有问题也都有多项式算法.

已发现有成千的组合最优化问题属于 NP 完全问题. 为求解该类中的问题，人们往往采用"启发式"方法. 这些方法一般地，不能保证求得问题的最优解，但常能得到较好的近似解.

网络最优化方法(network optimization method） 一类求解网络最优化问题的方法. 指对网络最优化问题应用网络拓扑或线图分析寻优的方法.

数学规划(mathematical programming） 运筹学的重要学科. 也是最优化的主要组成部分. 是研究在给定的条件下，求某一实值(或向量值)函数的极小化或极大化的学科. 极小化形式的数学规划问题表示为(MP)：

$$\begin{cases} \min\ f(\boldsymbol{x}), & (1) \\ \text{s.t.}\ g_i(\boldsymbol{x}) \leqslant 0 \quad (i=1,2,\cdots,p), & (2) \\ \quad h_j(\boldsymbol{x}) = 0 \quad (j=1,2,\cdots,q), & (3) \end{cases}$$

其中变量 $\boldsymbol{x}\in\mathbb{R}^n$($n$ 维欧氏空间，也可以是其他有关空间)，$f(\boldsymbol{x})$是 $\boldsymbol{x}$ 的实值(或向量值)函数，$g_i(\boldsymbol{x})(i=1,2,\cdots,p)$和 $h_j(\boldsymbol{x})(j=1,2,\cdots,q)$是 $\boldsymbol{x}$ 的实值函数，min 表示"极小化"，s. t. (sebject to)表示"受约束于". 上述(MP)中的 $\boldsymbol{x}$ 称为问题的决策变量，$f(\boldsymbol{x})$称为目标函数，式(2)和(3)称为约束条件. 换言之，数学规划研究如何求一满足约束条件(2)和(3)的 $\boldsymbol{x}^*\in\mathbb{R}^n$，使得目标函数值 $f(\boldsymbol{x}^*)$为最小(对于 $f(\boldsymbol{x})$是向量值函数的情况，其极小的意义另有定义). 按照决策变量、目标函数和约束条件的不同特点，数学规划分为(包括)：线性规划($f(\boldsymbol{x})$，$g_i(\boldsymbol{x})(i=1,2,\cdots,p)$和 $h_j(\boldsymbol{x})(j=1,2,\cdots,q)$都是 $\boldsymbol{x}$ 的线性函数)，非线性规划($f(\boldsymbol{x})$，$g_i(\boldsymbol{x})(i=1,2,\cdots,p)$和 $h_j(\boldsymbol{x})(j=1,2,\cdots,q)$中至少有一个是 $\boldsymbol{x}$ 的非线性函数)，整数规划($\boldsymbol{x}$ 的分量中全部或部分取整数值)，参数规划(在 $f(\boldsymbol{x})$，$g_i(\boldsymbol{x})(i=1,2,\cdots,p)$或 $h_j(\boldsymbol{x})(j=1,2,\cdots,q)$中含有某些参数)，多目标规划($f(\boldsymbol{x})$是一向量值函数)，动态规划(问题是多阶段性的)，随机规划($f(\boldsymbol{x})$，$g_i(\boldsymbol{x})(i=1,2,\cdots,p)$或 $h_j(\boldsymbol{x})(j=1,2,\cdots,q)$中含有随机变量)，以及组合规划(即组合最优化，由约束条件构成的集合是一有限集，$\boldsymbol{x}$ 是该集合的子集)等. 数学规划这一术语是 1950 年前后由多夫曼(Dorfman，R.)提出的. 1970 年成立国际数学规划协会(International Mathematical Programming Society)，从 1982 年开始，每次国际数学规划讨论会(3 年一次)上颁发两项奖金：丹齐克奖(数学规划)和富尔克森奖(组合数学).

决策变量(decision variable） 数学规划的基本概念之一. 即数学规划问题中的变量(参见"数学规划").

目标函数(objective function） 数学规划的基本概念之一. 在数学规划问题中，要求在一定条件下求极小化(或极大化)的函数(参见"数学规划"). 一般地，它是问题中变量的单值实值函数. 在多目标规划中，它是变量的向量值函数(参见"多目标规划").

约束条件(constraint condition） 数学规划的基本概念之一. 指数学规划问题中变量必须满足的条件. 用不等式表示的条件称为不等式约束条件，用等式表示的称为等式约束条件. 等式约束条件的数目不能超过变量的维数.

约束集(constraint set） 数学规划的基本概念之一. 指在数学规划问题中，满足所有约束条件的点组成的集合.

可行域(feasible region） 亦称允许域、能行域. 数学规划的基本概念之一. 即约束集(参见"约束集"). 一个数学规划问题的可行域可以是有界的，也可以是无界的.

可行解(feasible solution） 亦称可行点或允许解. 数学规划的基本概念之一. 指在数学规划问题中，满足所有约束条件的解(点).

最优解(optimal solution） 数学规划的基本概念之一. 指在数学规划问题中，使目标函数取最小值(对极大化问题取最大值)的可行解. 使目标函数取最小值的可行解称为极小解(点)，使其取最大值的可行解称为极大解(点). 极小解(点)或极大解(点)均称为最优解(点). 相应地，目标函数的最小值或最大值称为最优值. 有时，也将最优解(点)和最优值一起称为相应数学规划问题的最优解.

线 性 规 划

线性规划(linear programming） 数学规划中的较成熟的一个分支. 研究线性函数在满足线性等式和(或)不等式的约束条件下，达到最大值或最小值问题的学科. 其研究的问题称为线性规划问题，也简称线性规划，记为(LP).

线性规划的数学模型(mathematical model of

linear programming) 线性规划问题的一种数学表述.即:求一组 $x_j(j=1,2,\cdots,n)$,使满足

$$\begin{cases}\max(\min)\ c_1x_1+c_2x_2+\cdots+c_nx_n,\\ \text{s.t.}\ a_{11}x_1+a_{12}x_2+\cdots\\ \qquad +a_{1n}x_n\leqslant(\text{或}=,\text{或}\geqslant)b_1,\\ \quad a_{21}x_1+a_{22}x_2+\cdots\\ \qquad +a_{2n}x_n\leqslant(\text{或}=,\text{或}\geqslant)b_2,\\ \quad\cdots\cdots\cdots\cdots\\ \quad a_{m1}x_1+a_{m2}x_2+\cdots\\ \qquad +a_{mn}x_n\leqslant(\text{或}=,\text{或}\geqslant)b_m,\\ \quad x_1,x_2,\cdots,x_n\geqslant 0,\end{cases}$$

其中 $c_j,b_i,a_{ij}\ (i=1,2,\cdots,m;j=1,2,\cdots,n)$为常数.符号 s.t. 见"数学规划".

线性规划的标准型(standard form of linear programming) 线性规划模型的标准形式.其主要特征为:

1. 目标函数为极大化类型.
2. 所有的约束条件都是方程.
3. 所有约束方程右端的常数都是非负的.
4. 所有决策变量都是非负的.

其数学表示式为

$$\begin{cases}\max\ c_1x_1+c_2x_2+\cdots+c_nx_n,\\ \text{s.t.}\ a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1,\\ \quad a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2,\\ \quad\cdots\cdots\cdots\cdots\\ \quad a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m,\\ \quad x_j\geqslant 0\quad(j=1,2,\cdots,n),\end{cases}$$

其中 $b_i\geqslant 0\ (i=1,2,\cdots,m)$.上述表示式可缩写成

$$\begin{cases}\max\ \sum_{j=1}^{n}c_jx_j,\\ \text{s.t.}\ \sum_{j=1}^{n}a_{ij}x_j=b_i\ (i=1,2,\cdots,m),\\ \quad x_j\geqslant 0\qquad (j=1,2,\cdots,n),\end{cases}$$

其中 $b_i\geqslant 0\ (i=1,2,\cdots,m)$.线性规划标准型的矩阵形式为

$$\begin{cases}\max\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},\\ \text{s.t.}\ A\boldsymbol{x}=\boldsymbol{b},\\ \quad \boldsymbol{x}\geqslant\boldsymbol{0},\end{cases}$$

其中

$$\boldsymbol{c}=\begin{pmatrix}c_1\\c_2\\\vdots\\c_n\end{pmatrix},\quad A=\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{pmatrix},$$

$$\boldsymbol{x}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix},\quad \boldsymbol{b}=\begin{pmatrix}b_1\\b_2\\\vdots\\b_m\end{pmatrix},\quad \boldsymbol{0}=\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}.$$

称 $\boldsymbol{c}$ 为收益(或成本)系数向量,A 为约束方程组的消耗系数矩阵,$\boldsymbol{x}$ 为决策向量,$\boldsymbol{b}$ 为资源的限制向量.线性规划标准型的向量形式为

$$\begin{cases}\max\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},\\ \text{s.t.}\ \sum_{j=1}^{n}\boldsymbol{a}_jx_j=\boldsymbol{b},\\ \quad x_j\geqslant 0\quad(j=1,2,\cdots,n),\end{cases}$$

其中

$$\boldsymbol{a}_j=\begin{pmatrix}a_{1j}\\a_{2j}\\\vdots\\a_{mj}\end{pmatrix}\quad(j=1,2,\cdots,n).$$

松弛变量(slack variable) 线性规划问题中引入的变量.指当把线性规划问题化为标准型时,原不等式约束条件中引入的新的非负变量.当第 i 个约束条件为

$$a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n\leqslant b_i,$$

在不等式的左端加上变量 x_{n+i},使之成为

$$a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n+x_{n+i}=b_i,$$

则称 x_{n+i}为松弛变量.松弛变量始终保持非负值,它在最优解中的数值提供了极其有用的分析资料.

剩余变量(surplus variable) 线性规划问题中引入的变量.当不等式约束条件为

$$a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n\geqslant b_i,$$

在不等式的左端减去变量 x_{n+i},使成为

$$a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n-x_{n+i}=b_i,$$

则称 x_{n+i}为剩余变量.剩余变量取非负值.

基矩阵(basic matrix) 线性规划问题中的矩阵.指在线性规划问题中,消耗系数矩阵 A 的 m 阶非奇异子矩阵,用 B 表示,即 $|B|\neq 0$.其中 A 是约束方程组的 $m\times n$ 阶系数矩阵(参见"线性规划问题的标准型"),秩 $r(A)=m\leqslant n$.基矩阵简称基.

基向量(basic vector) 线性规划问题中的一种向量.若基矩阵 B 可表示成 $B=(\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_m)$,其中 $\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_m$ 是系数矩阵 A 的 m 个列向量,称 $\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_m$ 为基矩阵 B 对应的基向量.

基变量(basic variable) 线性规划中的一种变量.即与基向量 $\boldsymbol{p}_j$ 对应的变量 $x_j(j=1,2,\cdots,m)$.

非基变量(nonbasic variable) 线性规划中的一种变量.即除基变量外的其余 $n-m$ 个变量.

基本解(basic solution) 线性规划的重要概念.非基变量为零时约束方程组的解称为对应于基 B 的一个基本解.有一个基,就有一个基本解与之对应.线性规划问题的基本解的个数是有限的,它不超过 C_n^m 个.由于$r(A)=m\leqslant n$,所以每一个基本解非零分量的个数不超过 m.

基本可行解(basic feasible solution) 线性规

划的重要概念. 在线性规划问题中,满足非负约束条件的基本解,称基本可行解,简称基可行解. 线性规划问题如果有可行解,则必有基可行解. 可行解是基可行解的充分必要条件为:它的非零分量所对应的系数矩阵列向量是线性无关的.

可行基(feasible basis)　线性规划的重要概念. 指相应于基可行解的基.

基本最优解(basic optimal solution)　线性规划的重要概念. 指线性规划问题中使目标函数达到最优值的基可行解.

最优基(optimal basis)　线性规划的重要概念. 指相应于基本最优解的基.

图解法(diagram method)　求解仅含两个变量的线性规划问题的一种方法. 只含两个变量的线性规划问题,由约束条件确定的可行域可以在二维平面上表示出来. 按照一定规则,在可行域上移动目标函数的等值线,从而得到线性规划问题的最优解. 这里的可行域是凸区域,最优解必在可行域的某个顶点上达到(见图).

凸集(convex set)　具某种特性的点集. 设 S 是线性空间中的集合,若连结 S 中任意两点的线段仍在 S 中,则称 S 是线性空间中的凸集. 或者说,若对任意实数 $\alpha\in(0,1)$ 和任意两点 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$,恒有 $\boldsymbol{x}=\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2\in S$,则称 S 为凸集. 线性规划问题的可行域 $S=\{\boldsymbol{x}\mid A\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geqslant\boldsymbol{0}\}$ 是凸集.

凸组合(convex combination)　若干个点的某种特定意义下的非负线性组合. 设 $\boldsymbol{x}=\alpha_1\boldsymbol{x}^1+\alpha_2\boldsymbol{x}^2+\cdots+\alpha_k\boldsymbol{x}^k$,其中 $\boldsymbol{x}^1,\boldsymbol{x}^2,\cdots,\boldsymbol{x}^k$ 是线性空间中的 k 个点,$0\leqslant\alpha_i\leqslant1(i=1,2,\cdots,k)$,且

$$\sum_{i=1}^{k}\alpha_i=1,$$

则称 $\boldsymbol{x}$ 为 $\boldsymbol{x}^1,\boldsymbol{x}^2,\cdots,\boldsymbol{x}^k$ 的凸组合.

极点(extreme point)　线性规划的重要概念. 设 S 是凸集,点 $\boldsymbol{x}\in S$. 若 $\boldsymbol{x}$ 不能成为 S 中任何两点连线的内点,则称它为 S 的极点. 或者说,不存在两个不同的点 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$,使 $\boldsymbol{x}=\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2\in S(0<\alpha<1)$,则称 $\boldsymbol{x}$ 为 S 的极点(或顶点). 有界凸集中任何一点都可表示为 S 的极点的凸组合. 线性规划问题的可行域如果有界,则最优值必在 S 的极点上达到;且可行域中的点 $\boldsymbol{x}$ 是极点的充分必要条件为:点 $\boldsymbol{x}$ 是基可行解.

单纯形(simplex)　一种几何形体. 在 n 维空间中,由不处于同一 $(n-1)$ 维超平面上的 $(n+1)$ 个点生成的凸多面体. 例如,一维空间中的线段、二维空间中的三角形、三维空间中的四面体等几何形体都是单纯形.

线性规划的典式(typical form of linear programming)　线性规划的标准型的典型形式. 即:给定一组可行基后,可将线性规划问题的目标函数和约束方程组变换成与之等价的下列形式:

$$\begin{cases}\max f=f_0+\sum\limits_{j=m+1}^{n}\lambda_j x_j,\\ \text{s.t. } x_1=\bar{b}_1-\sum\limits_{j=m+1}^{n}\bar{a}_{1j}x_j,\\ \qquad x_2=\bar{b}_2-\sum\limits_{j=m+1}^{n}\bar{a}_{2j}x_j,\\ \qquad \cdots\cdots\cdots\cdots\\ \qquad x_m=\bar{b}_m-\sum\limits_{j=m+1}^{n}\bar{a}_{mj}x_j,\\ \qquad x_j\geqslant0\ (j=1,2,\cdots,n),\end{cases}$$

其中 $x_1,x_2,\cdots,x_m$ 为基变量,$x_{m+1},x_{m+2},\cdots,x_n$ 为非基变量. 线性规划的这种形式称为用非基变量分别表示基变量和目标函数的典式.

单纯形表(simplex tableau)　利用单纯形法求解线性规划问题过程中的数据表. 为了计算方便,单纯形法的迭代计算可以在一张数据表上进行,这张数据表被称为对应于初始可行基 B 的单纯形表,记为 $T(B)$. 此表是这样得到的:依据线性规划的典式,把目标函数改写为

$$-f+\sum_{j=1}^{n}\lambda_j x_j=-f_0\ \ (\lambda_1=\lambda_2=\cdots=\lambda_m=0),$$

并令 $b_{ij}=\bar{a}_{ij},b_{i0}=\bar{b}_i(i=1,2,\cdots,m;j=m+1,\cdots,n)$,于是有新的方程组,把这个方程组的系数与常数填入表中,便得如下的单纯形表 $T(B)$:

$c_j\to$			c_1	c_2	$\cdots$	c_m	c_{m+1}	c_{m+2}	$\cdots$	c_n	θ_i
$\boldsymbol{c}_B$	$\boldsymbol{x}_B$	$\bar{\boldsymbol{b}}$	x_1	x_2	$\cdots$	x_m	x_{m+1}	x_{m+2}	$\cdots$	x_n	
c_1	x_1	b_{10}	1	0	$\cdots$	0	b_{1m+1}	b_{1m+2}	$\cdots$	b_{1n}	
c_2	x_2	b_{20}	0	1	$\cdots$	0	b_{2m+1}	b_{2m+2}	$\cdots$	b_{2n}	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
c_m	x_m	b_{m0}	0	0	$\cdots$	1	b_{mm+1}	b_{mm+2}	$\cdots$	b_{mn}	
	$-f_0$	$-f_0$	0	0	$\cdots$	0	λ_{m+1}	λ_{m+2}	$\cdots$	λ_n	$\leftarrow\lambda_j$

单纯形法(simplex method)　求解线性规划问题的基本方法. 此方法是丹齐克(Dantzig,G. B.)于1947年提出来的. 方法的基本思路是:根据线性规划问题的标准型形式,从可行域中一个基可行解开始,变换到另一个基可行解,并且使目标函数的值逐步增大;当目标函数达到最大值时,就得到了该线性规划问题的最优解. 由于线性规划问题仅有有限个基可行解,所以如不出现循环,全部迭代过程可在有限次内终止. 此时,或者已得到问题的最优解,或者判定问题无有限最优解.

换基迭代(basis iteration)　求解线性规划问题过程中的迭代技巧. 是从一个基可行解到另一个基可行解的迭代. 在单纯形表(参见"单纯形表")上的换基迭代过程是:

1. 确定入基变量. 若在 $T(B)$ 的检验数 $\lambda_j(j=1,2,\cdots,n)$ 中有检验数 $\lambda_s>0$, 且 λ_s 所在列的其他元素中有 $b_{is}>0(1\leqslant i\leqslant m)$, 则取 $s=\min\limits_j\{j\mid\lambda_j>0\}$, 即 $T(B)$ 中最左边的一个正检验数 λ_s(或取 $\max\limits_j\{\lambda_j\mid\lambda_j>0\}=\lambda_s$), 让其对应的变量 x_s 为入基变量.

2. 求主元, 确定出基变量. 按最小比值原则

$$\theta=\min\{\theta_i\}=\min\left\{\frac{b_{i0}}{b_{is}}\middle| b_{is}>0\right\}$$
$$=\frac{b_{r0}}{b_{rs}}\ (1\leqslant r\leqslant m),$$

其中 b_{rs} 为主元, 记为 b_{rs}^*. 主元 b_{rs} 所在行的基变量 x_r 就是要确定的出基变量.

3. 以 b_{rs} 为主元, 进行初等行变换. 将入基变量 x_s 所在的列变为单位向量, 即 $(b_{1s},\cdots,b_{r-1\,s},b_{rs},b_{r+1\,s},\cdots,b_{ms},\lambda_s)^{\mathrm T}\to(0,\cdots,0,1,0,\cdots,0,0)^{\mathrm T}$. 与此同时, 原 $T(B)$ 中的各元素按以下各式计算, 变为相应的新元素:

$$\bar b_{rj}=\frac{b_{rj}}{b_{rs}}\ (j=0,1,\cdots,n),$$

$$\bar b_{ij}=b_{ij}-\frac{b_{rj}}{b_{rs}}b_{is}$$
$$(i=1,2,\cdots,m,i\neq r,j=0,1,\cdots,n),$$

$$-\bar f_0=-f_0-\frac{b_{r0}}{b_{rs}}\lambda_s,$$

$$\bar\lambda_j=\lambda_j-\frac{b_{rj}}{b_{rs}}\lambda_s\ (j=1,2,\cdots,n),$$

于是得到新基 $\bar B=(\boldsymbol p_1,\boldsymbol p_2,\cdots,\boldsymbol p_{r-1},\boldsymbol p_s,\boldsymbol p_{r+1},\cdots,\boldsymbol p_m)$ 所对应的单纯形表 $T(\bar B)$:

$c_j\to$		c_1	$\cdots$	c_r	$\cdots$	c_m	c_{m+1}	$\cdots$	c_s	$\cdots$	c_n
$\boldsymbol x_B$	$\bar{\boldsymbol b}$	x_1	$\cdots$	x_r	$\cdots$	x_m	x_{m+1}	$\cdots$	x_s	$\cdots$	x_n
x_1	$\bar b_{10}$	1	$\cdots$	$-\frac{b_{1s}}{b_{rs}}$	$\cdots$	0	$\bar b_{1m+1}$	$\cdots$	0	$\cdots$	$\bar b_{1n}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_{r-1}	$\bar b_{r-1\,0}$	$\vdots$		$\vdots$		$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_s	$\bar b_{r0}$	0	$\cdots$	$\frac{1}{b_{rs}}$	$\cdots$	0	$\bar b_{rm+1}$	$\cdots$	1	$\cdots$	$\bar b_{rn}$
x_{r+1}	$\bar b_{r+1\,0}$	$\vdots$		$\vdots$		$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_m	$\bar b_{m0}$	0	$\cdots$	$-\frac{b_{ms}}{b_{rs}}$	$\cdots$	1	$\bar b_{mm+1}$	$\cdots$	0	$\cdots$	$\bar b_{mn}$
$-f_0$	$-\bar f_0$	0	$\cdots$	$\bar\lambda_r$	$\cdots$	0	$\bar\lambda_{m+1}$	$\cdots$	0	$\cdots$	$\bar\lambda_n$

单纯形法的主要步骤:

1. 确定初始基可行解.

2. 建立初始可行基的单纯形表(参见"单纯形表").

3. 判别最优解. 若所有的检验数 $\lambda_j\leqslant 0(j=1,2,\cdots,n)$, 则单纯形表中所确定的基可行解就是基本最优解; 若有一检验数 $\lambda_s>0(m+1\leqslant s\leqslant n)$, 且所在列的其他各数 $b_{is}\leqslant 0(i=1,2,\cdots,m)$, 则无有限最优解.

4. 换基迭代(参见"换基迭代").

反复步骤 3,4, 直至获得基本最优解, 或判定无有限最优解, 计算终止.

非退化的基可行解(nondegenerate basic feasible solution)　线性规划的基本概念之一. 指所有基变量均取正值的基可行解.

退化的基可行解(degenerate basic feasible solution)　线性规划的基本概念之一. 指至少有一个基变量的值等于零的基可行解.

退化的线性规划问题(degenerate linear programming problem)　一类线性规划问题. 是具有至少一组退化的基可行解的线性规划问题.

摄动法(perturbation method)　避免单纯形法在求解线性规划问题过程中出现循环的一种方法. 查恩斯(Charnes, A.)于 1952 年提出摄动法, 基本思路是: 设 $B=(\boldsymbol p_1,\boldsymbol p_2,\cdots,\boldsymbol p_m)$ 和 $\boldsymbol x_B$ 分别是线性规划问题的一个可行基和基可行解, 则有 $\boldsymbol p_1x_1+\boldsymbol p_2x_2+\cdots+\boldsymbol p_mx_m=\boldsymbol b$. 如果 $\boldsymbol x_B$ 是退化的基可行解, 它的基变量 $x_i(i=1,2,\cdots,m)$ 中就有等于零的. 若对 $\boldsymbol b$ 能作一个微小的变动, 使系数 $\boldsymbol p_j(j=1,2,\cdots,m)$ 全部是正的, 则 $\boldsymbol x_B$ 就变成非退化的. 若变动 $\boldsymbol b$ 使得所有的可行基都具有上述性质, 则这个变动 $\boldsymbol b$ 以后的线性规划问题就是非退化的, 从而可以用单纯形法在有限次迭代后得到解, 然后再把 $\boldsymbol b$ 变回来, 得到原问题的最优解. 用单纯形法解线性规划问题的第一个循环例子是霍夫曼(Hoffman, A. J.)提出的. 以后, 贝尔(Beale, E. M. L.)于 1955 年又提出了一个循环的例子.

摄动问题(perturbation problem)　一类线性规划问题. 指对线性规划问题的约束条件进行微小变动后得到的新的线性规划问题. 即在常数项 b_i 后面加上 $a_{i1}\varepsilon+a_{i2}\varepsilon^2+\cdots+a_{in}\varepsilon^n$, 得到一个新的线性规划问题. 这里 ε 是一个充分小的正数, ε^j 表示 ε 的 j 次方. 利用单纯形法求解, 不会出现循环, 并且不用找出 ε, 只要让 ε 充分小就可以了. 在 ε 充分小时, 令任一基可行解中的 ε 等于零, 就得到原线性规划问题的一个基可行解. 若没有可行解, 则原线性规划问题也没有可行解. 在迭代过程中, ε^j 的系数在单纯形表上都出现, 不用单独找; 选择出基变量也可以在单纯表上进行. 找初始基可行解的方法是: 先找原线性规划问题的初始基可行解, 然后把变量的下标换一

下即可.

字典序法(dictionary ordered method) 一种摄动法.由丹齐克(Dantzig,G.B.)于1954年提出,格莫理(Gomory,R.E.)在关于解整数线性规划问题的文章中又提到字典序法,且叙述得更加完整.设$\boldsymbol{v}=(v_1,v_2,\cdots,v_n)$是$n$维空间中一个向量.若$\boldsymbol{v}$的第一个非零分量是正的,则称它是字典式正的,记为$\boldsymbol{v}\underset{L}{>}\boldsymbol{0}$.若两个向量$\boldsymbol{u},\boldsymbol{v}$,有$\boldsymbol{v}-\boldsymbol{u}\underset{L}{>}\boldsymbol{0}$,则称向量$\boldsymbol{v}$字典式大于$\boldsymbol{u}$.关于向量序列$\{\boldsymbol{v}^t\}$,若对所有的$t$都有$\boldsymbol{v}^{t+1}-\boldsymbol{v}^t\underset{L}{>}\boldsymbol{0}$,则称$\{\boldsymbol{v}^t\}$是字典式增加的.若一个向量$\boldsymbol{v}$满足$-\boldsymbol{v}\underset{L}{>}\boldsymbol{0}$,则称$\boldsymbol{v}$是字典式负的.记为$\boldsymbol{v}\underset{L}{<}\boldsymbol{0}$.若$\boldsymbol{v}-\boldsymbol{u}\underset{L}{<}\boldsymbol{0}$,则称$\boldsymbol{v}$字典式小于$\boldsymbol{u}$.记号$\boldsymbol{v}\underset{L}{\geqslant}\boldsymbol{0}$,表示$\boldsymbol{v}=\boldsymbol{0}$或$\boldsymbol{v}\underset{L}{>}\boldsymbol{0}$.在单纯形法中,用$\boldsymbol{v}_i$表示单纯形表中与基变量$x_i$对应的行的系数向量.出基变量的选择法则为:设$\lambda_k>0$,选定$x_k$是入基变量,令$S_k=\{i\mid b_{ik}>0,i\in S\}$,其中$S$为基变量指标集合,作$\boldsymbol{u}_i=\boldsymbol{v}_i/b_{ik}(i\in S_k)$.若$\boldsymbol{u}_r$是这些向量$\boldsymbol{u}_i$中字典式最小的,即

$$\boldsymbol{u}_r=\operatorname{lex}\min_{i\in S_k}\boldsymbol{u}_i\text{(lex min 代表字典式最小)},$$

则选x_r为出基变量.这样选择出基变量是不会产生循环的(参见"换基迭代").

布兰德规则(Bland rule) 一种用单纯形法求解线性规划问题时避免循环的方法.它是布兰德(Bland,R.G.)于1977年提出的.此方法比字典序法简单得多,在国际上受到很多人的重视,认为是线性规划中一项很好的成果,更有利于在计算机上施行.布兰德规则是:在换基迭代中(参见"换基迭代"),选取与下标最小的(即最左边的)正检验数λ_s相应的非基变量x_s为入基变量,其中$s=\min\{j\mid\lambda_j>0\}$;且当出现两个以上相同的最小比值$\theta_i$时,选取下标最小的$\theta_r$相应的基变量$x_r$为出基变量,其中

$$r=\min\left\{t\,\middle|\,\theta_t=\min\left\{\frac{b_{i0}}{b_{is}}\,\middle|\,b_{is}>0,1\leqslant i\leqslant m\right\}\right\}.$$

人工变量(artificial variable) 亦称人造变量.求解线性规划问题时人为加入的变量.用单纯形法求解线性规划问题,都是在具有初始可行基的条件下进行的,但约束方程组的系数矩阵A中所含的单位向量常常不足m个,此时可加入若干(至多m)个新变量,称这些新变量为人工变量.如设(LP)为

$$\begin{cases}\max\ c_1x_1+c_2x_2+\cdots+c_nx_n,\\ \text{s.t.}\ a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1,\\ \qquad a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2,\\ \qquad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \qquad a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m,\\ \qquad x_j\geqslant 0\quad(j=1,2,\cdots,n),\end{cases}\tag{1}$$

其中$b_i\geqslant 0(i=1,2,\cdots,m)$.引入新变量将(1)变为

$$\begin{cases}y_1+a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1,\\ y_2+a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2,\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ y_m+a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m,\\ x_j\geqslant 0,y_i\geqslant 0\ (i=1,2,\cdots,m;j=1,2,\cdots,n).\end{cases}\tag{2}$$

称新变量$y_1,y_2,\cdots,y_m$为人工变量.

约束条件(2)中的$B=(\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_m)$称为人工基.原约束条件(1)中不包含人工变量$y_1,y_2,\cdots,y_m$.方程组(1)和(2)是不等价的,仅当所有的人工变量$y_1,y_2,\cdots,y_m$都等于零时,它们才等价.因此,必须在换基迭代中逐步将人工变量替换成非基变量,或使其值为零.

大M法(big M method) 寻找线性规划问题初始基可行解的一种方法.在线性规划问题的约束条件中加入人工变量后,要求在目标函数中相应地加入以M或$-M$为系数的项.在极小化问题中,对人工变量赋于一个很大的正数M作为其系数;在极大化问题中,对人工变量赋于$-M$作为其系数.应用单纯形法在改进目标函数的过程中,如果原问题存在最优解,必然使人工变量逐步变为非基变量,或使其值为零.否则,目标函数值将不可能达到最小或最大.在迭代过程中,若全部人工变量变成非基变量,则可把人工变量所在的列从单纯形表中删去,此时便找到原问题的一个初始基可行解.若此基可行解不是原问题的最优解,则继续迭代,直至求得最优解为止.

两阶段法(two-phase method) 寻找线性规划问题初始基可行解的一种方法.把增加人工变量的线性规划问题分为两个阶段去求解.第一阶段是构造一个辅助的人工目标函数,即

$$\min Z=\sum_{i=1}^{m}y_i\ \text{或}\ \max Z'=\sum_{i=1}^{m}(-y_i).$$

若原问题有可行解,则在本阶段的最终单纯形表中,必有$Z=0$和$y_i=0(i=1,2,\cdots,m)$,并使人工变量均为非基变量.此时,划去人工变量所在的列与人工目标函数所在的行,就得到原问题的初始可行基对应的单纯形表,进入第二阶段.否则,在第一阶段的某张表中必有$Z\neq 0$,且人工变量中至少还有一个不为零值的基变量.此时表明原问题无可行解,从而不必进行第二阶段,计算终止.第二阶段是在第一阶段得到的初始可行基对应的单纯形表的基础上,按原问题的目标函数进行换基迭代,直至求得原问题的最优解或判定无有限最优解为止.有时在第一阶段的最终表中,有可能出现$Z=0$,且在基变量中仍然含有一个或几个取值为零的人工变量(退化解)的情况.虽然现行基不可作为原问题的初始可行基,但此时,可继续进行一次或几次换基迭代,将基变量中的

人工变量逐步替换为非基变量，而让不是人工变量的非基变量成为基变量. 由于迭代后的入基变量的取值仍为零，所以并不影响人工目标函数值. 直至所有的人工变量均为非基变量时，才得到原问题的初始基可行解，从而转入第二阶段.

大 M 法与两阶段法都是在原问题缺少初始可行基的情况下利用引入人工变量构造人工基，以达到运用单纯形法求解原问题的目的. 但由于大 M 法引进的 M 必须是个很大的正的常数，而决策变量、松弛变量的系数与它相比显得很小，这在计算机中计算时很容易产生误差，从而使求解的最终结果与原问题真正的最优解不一致. 在大规模线性规划问题的求解中(参见"大规模线性规划问题")，通常采用两阶段法运算.

单纯形表的矩阵形式(matrix form of simplex tableau)　对应于可行基的单纯形表的矩阵形式. (LP)的数学模型的矩阵形式为

$$\begin{pmatrix} 0 & I & B^{-1}N & B^{-1} \\ 1 & 0 & c_N - c_B B^{-1}N & -c_B B^{-1} \end{pmatrix} \begin{pmatrix} -f \\ x_B \\ x_N \\ x_L \end{pmatrix} = \begin{pmatrix} B^{-1}b \\ -c_B B^{-1}b \end{pmatrix},$$

其中 f 是目标函数，B 是基矩阵，N 是非基矩阵，x_B 是基变量，x_N 是非基变量，$x_L=(x_{n+1},x_{n+2},\cdots,x_{n+m})^T$ 是松弛变量，I 是 m 阶单位矩阵. 对应于可行基 B 的单纯形表 $T(B)$ 的矩阵形式(参见"单纯形表")为

$c_j \to$	c_B	c_N	c_L
	x_B	x_N	x_L
$B^{-1}b$	I	$B^{-1}N$	B^{-1}
$-c_B B^{-1}b$	0	$c_N - c_B B^{-1}N$	$-c_B B^{-1}$

改进单纯形法(revised simplex method)　一种单纯形方法. 只对单纯形表中非基变量的检验数、入基变量所在列、以及常数列等有关数据进行计算，而其余的各列元素都不计算，以减少计算量和误差的单纯形法(参见"单纯形法"和"单纯形表"). 对于极大化的(LP)，其步骤如下：

1. 引入松弛变量和人工变量，把约束方程组化成典型方程组. 设初始基矩阵为 B，置 $B^{-1}=B=I$，$\bar{b}=b$，并相应地确定基变量.

2. 计算单纯形乘子 $\pi=c_B B^{-1}$.

3. 按顺序逐个计算非基变量的检验数 $\lambda_j=c_j-\pi^T p_j$. 若所有的 $\lambda_j\leqslant 0$，则求得的解为最优解，并计算最优值 $s^*=\pi^T b$. 否则，找出第一个为正数的检验数 λ_s，确定其对应的非基变量 x_s 为入基变量.

4. 计算 $\bar{p}_s=B^{-1}p_s$. 若 $B^{-1}p_s\leqslant 0$，则无有限最优解. 否则，进行下一步.

5. 由 $\bar{b}$ 和 $\bar{p}_s$，按最小比值规则

$$\theta=\min\{\theta_i\}=\min\left\{\frac{b_{i0}}{b_{is}}\,\middle|\, b_{is}>0\right\}=\frac{b_{r0}}{b_{rs}},$$

确定 b_{rs} 为主元，其对应的基变量 x_r 为出基变量. 若同时有几个最小的 θ_i，则取 $r=\min\{i\,|\,\theta_i\}$.

6. 以 b_{rs} 为主元，对 B^{-1} 和 $\bar{p}_s$ 实施初等行变换，得到新的基矩阵的逆矩阵 $\bar{B}^{-1}$，并计算 $\bar{b}=\bar{B}^{-1}b$，置 $B^{-1}=\bar{B}^{-1}$，返回步骤 2.

多项式算法(polynomial algorithm)　一类求解最优化问题的重要算法. 指基本运算次数可控制在问题规模的多项式函数之内的算法. 也即设 n 表示问题的规模(决策变量数、约束条件数、输入数码长度)，$p(n)$ 表示对算法影响最坏的情况下所需的运算次数. 若 $p(n)$ 是 n 的多项式函数，则称该算法是多项式算法或多项式时间算法. 多项式算法这一概念是 20 世纪 60 年代提出来的.

n 维椭球(n-dimensional ellipsoid)　n 维空间中满足特定条件的点集. 即所有满足 $(x-x^0)^T Q^{-1}(x-x^0)\leqslant 1$ 的点 $x\in \mathrm{R}^n$ 的集合，其中 Q 是一个 n 阶对称正定阵，x^0 是一个给定的 n 维列向量. 用 $E(x^0,Q)$ 表示由 x^0 和矩阵 Q 决定的椭球. n 维椭球的体积为 $V=k(n)\sqrt{|Q|}$，其中 $k(n)$ 是一个随维数 n 变化的数，矩阵

$$Q=\begin{pmatrix} a_1^2 & 0 & \cdots & 0 \\ 0 & a_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n^2 \end{pmatrix},$$

Q 的行列式为 $|Q|=a_1^2 a_2^2\cdots a_n^2$.

椭球算法(ellipsoid algorithm)　一种求解线性规划问题的多项式算法(参见"多项式算法"和"哈奇扬算法"). 椭球算法是由朱丁(Judin, D. B.)和内密洛夫斯基(Nemirovskii, A. S.)在前人研究方法的基础上发展起来的，主要用来求解凸规划问题. 后来，哈奇扬(Хачиян, Л. Г.)用它来解具有整系数的线性不等式组，提出了求解线性规划问题的多项式算法.

哈奇扬算法(Khachian algorithm)　亦称椭球算法. 一种求解线性规划问题的多项式算法(参见"多项式算法"). 也即解"严格"整系数线性不等式组

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n<b_1, \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n<b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n<b_m, \end{cases}$$

(a_{ij} 与 b_i 都是整数)的一种算法. 求解线性规划问题的最优解，可以归结为解上述线性不等式组. 哈奇扬

算法是一种迭代法，每迭代一次就以某一点 x^i 为中心，并依照一定规则构造一个椭球 E_i（参见“n 维椭球”）．第一个椭球 E_1 就是圆 $\|x\|<2^L$，它的中心 x^1 是原点．迭代过程就是从 $\{x^1,Q_1\}$ 得到 $\{x^2,Q_2\}$，再到 $\{x^3,Q_3\}$，…，最多迭代 $6n^2L$ 次．这里，$Q_1=2^{2L}I$，

$$Q_{i+1}=\frac{n^2}{n^2-1}\left[Q_i-\frac{2}{n+1}\cdot\frac{(Q_iA_r^{\mathrm T})(Q_iA_r^{\mathrm T})^{\mathrm T}}{(Q_iA_r^{\mathrm T})A_rQ_iA_r^{\mathrm T}}\right],$$

L 是输入长度，n 是未知数的个数，A_r 是由 $A_rx^i\geqslant b_r$ 确定的，$(Q_iA_r^{\mathrm T})$ 是一个列向量．通过一系列椭球的迭代得到最优解．这一算法是苏联数学家哈奇扬（Хачиян，Л. Г.）于 1979 年提出的，1981 年作了完整的证明．这一算法的重要性在于，他第一个证明了线性规划问题是存在多项式算法的．1982 年 8 月在第 11 届国际数学规划讨论会上，哈奇扬的论文《线性规划的一个多项式算法》获得了富尔克森奖．

卡马卡算法（Karmarkar algorithm） 一种求解线性规划问题的新的多项式算法．它是继哈奇扬算法之后的又一种求解线性规划问题的多项式算法（参见“多项式算法”）．这种算法的基本思路是：对线性规划问题

$$\begin{cases}\min\ \boldsymbol{c}^{\mathrm T}\boldsymbol{x},\\ \text{s.t.}\ \ A\boldsymbol{x}=\boldsymbol{b},\\ \qquad \boldsymbol{x}\geqslant\boldsymbol{0},\end{cases}$$

先定义一个势函数

$$f(\boldsymbol{x})=n\lg\boldsymbol{c}^{\mathrm T}\boldsymbol{x}-\sum_{j=1}^{n}\lg x_j,$$

其中 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm T}$；然后对势函数进行投影变换，并利用梯度投影方法使势函数的值下降，通过一系列这种运算，逐步逼近最优解．这种算法也是一种迭代法，是由美籍印度数学家卡马卡（Karmarkar，N.）于 1984 年提出的．

对偶线性规划

对偶线性规划（dual linear programming） 与一线性规划（问题）相关的另一个线性规划（问题）．考虑线性规划问题（LP）（参见“线性规划的数学模型”和“线性规划的标准型”）：

$$\begin{cases}\max\ \boldsymbol{c}^{\mathrm T}\boldsymbol{x},\\ \text{s.t.}\ \ A\boldsymbol{x}\leqslant\boldsymbol{b},\\ \qquad \boldsymbol{x}\geqslant\boldsymbol{0},\end{cases}$$

以及与它相关的另一个线性规划问题（DLP）：

$$\begin{cases}\min\ \boldsymbol{y}^{\mathrm T}\boldsymbol{b},\\ \text{s.t.}\ \ \boldsymbol{y}^{\mathrm T}A\geqslant\boldsymbol{c},\\ \qquad \boldsymbol{y}\geqslant\boldsymbol{0}.\end{cases}$$

称前一个问题为原规划（问题），后一个问题为前一个问题的对偶线性规划（问题）．它们互为对偶线性规划问题．对偶规划最初是由冯·诺伊曼（von Neumann，J.）于 1947 年提出来的，以后库恩（Kuhn，H. W.）和塔克尔（Tucker，A. W.）证明了对偶定理．哥德曼（Goldman，A. J.）和塔克尔于 1956 年比较系统地叙述了对偶规划的理论．

对偶线性规划的经济背景是：若原问题是利用有限资源安排最优生产方案，以获得最大总产值的线性规划问题，则它的对偶问题就是在相同资源的条件下，正确估计资源的使用价值，以达到支付最少费用的线性规划问题．简言之，若原问题为求解资源的最优配置问题，则对偶问题就是求解估价资源的使用价值问题．

对称形式的对偶线性规划（dual linear programming of symmetric form） 一类对偶线性规划问题（参见“对偶线性规划”）．一对对称形式的对偶线性规划的特征是：

1．一个问题求极大，另一个问题求极小．

2．求极大问题中的约束条件为“$\leqslant$”，求极小问题中的约束条件为“$\geqslant$”．

3．原问题中有 n 个变量，对偶问题中就有 n 个约束条件．原问题中有 m 个约束条件，对偶问题中就有 m 个变量（称为对偶变量）．

4．原问题的目标函数中变量的系数就是对偶问题约束条件的常数项．原问题中约束条件的常数项就是对偶问题目标函数中变量的系数．

5．两个互为对偶问题的约束条件的系数矩阵互为转置矩阵．

6．原问题与对偶问题中的变量均有非负约束．

对偶变量（dual variable） 对偶线性规划问题中的变量（参见“对称形式的对偶线性规划”）．

非对称形式的对偶线性规划（dual linear programming of asymmetric form） 一类对偶线性规划（问题）．考虑线性规划问题

$$\begin{cases}\max\ \boldsymbol{c}^{\mathrm T}\boldsymbol{x},\\ \text{s.t.}\ \ A\boldsymbol{x}=\boldsymbol{b},\\ \qquad \boldsymbol{x}\geqslant\boldsymbol{0},\end{cases}\tag{1}$$

以及与它相关的另一个线性规划问题

$$\begin{cases}\min\ \boldsymbol{y}^{\mathrm T}\boldsymbol{b},\\ \text{s.t.}\ \ \boldsymbol{y}^{\mathrm T}A\geqslant\boldsymbol{c},\\ \qquad \boldsymbol{y}\ \text{无非负约束}.\end{cases}\tag{2}$$

称（2）是（1）的非对称形式的对偶线性规划（问题）．若原问题的约束条件是等式约束，则它的对偶变量就无非负限制．反之，若在线性规划的原问题中，决策变量无非负约束，则它的对偶问题对应于该变量的约束条件应取等式．

混合型对偶线性规划（mixed type dual linear programming） 一类对偶线性规划（问题）．称

$$\begin{cases}\max\ \boldsymbol{c}^{(1)^{\mathrm{T}}}\boldsymbol{x}^{(1)}+\boldsymbol{c}^{(2)^{\mathrm{T}}}\boldsymbol{x}^{(2)},\\ \text{s.t. } A_{11}\boldsymbol{x}^{(1)}+A_{12}\boldsymbol{x}^{(2)}\leqslant\boldsymbol{b}^{(1)},\\ \quad A_{21}\boldsymbol{x}^{(1)}+A_{22}\boldsymbol{x}^{(2)}=\boldsymbol{b}^{(2)},\\ \quad \boldsymbol{x}^{(1)}\geqslant\boldsymbol{0},\boldsymbol{x}^{(2)}\text{ 无非负约束}\end{cases}$$

与

$$\begin{cases}\min\ \boldsymbol{y}^{(1)^{\mathrm{T}}}\boldsymbol{b}^{(1)}+\boldsymbol{y}^{(2)^{\mathrm{T}}}\boldsymbol{b}^{(2)},\\ \text{s.t. } \boldsymbol{y}^{(1)^{\mathrm{T}}}A_{11}+\boldsymbol{y}^{(2)^{\mathrm{T}}}A_{21}\geqslant\boldsymbol{c}^{(1)},\\ \quad \boldsymbol{y}^{(1)^{\mathrm{T}}}A_{12}+\boldsymbol{y}^{(2)^{\mathrm{T}}}A_{22}=\boldsymbol{c}^{(2)},\\ \quad \boldsymbol{y}^{(1)}\geqslant\boldsymbol{0},\boldsymbol{y}^{(2)}\text{ 无非负约束}\end{cases}$$

是互为混合型对偶线性规划(问题).

对称性(symmetry) 对偶规划问题的一个重要性质. 即对偶问题的对偶是原问题.

弱对偶性(weak duality) 对偶线性规划问题的一个基本性质. 即:$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^{(0)}\leqslant\boldsymbol{y}^{(0)^{\mathrm{T}}}\boldsymbol{b}$,其中 $\boldsymbol{x}^{(0)}$ 和 $\boldsymbol{y}^{(0)}$ 分别是原问题和对偶问题的可行解. 从它可以推出三条性质:

1. 原问题的任意一个可行解对应的目标函数值是对偶问题目标函数值的一个下界;对偶问题的任意一个可行解对应的目标函数值是原问题的目标函数值的一个上界.

2. 若原问题可行,但其目标函数值无上界,则对偶问题不可行;若对偶问题可行,但其目标函数值无下界,则原问题不可行.

3. 若原问题可行,而对偶问题不可行,则原问题无上界;若对偶问题可行,而原问题不可行,则对偶问题无下界.

最优准则(optimal criterion) 判别最优解的一个标准. 设 $\boldsymbol{x}^*$ 和 $\boldsymbol{y}^*$ 分别是原问题和对偶问题的可行解(参见"对偶线性规划"和"可行解"),且 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^*=\boldsymbol{y}^{*^{\mathrm{T}}}\boldsymbol{b}$,则 $\boldsymbol{x}^*$ 和 $\boldsymbol{y}^*$ 分别是它们的最优解.

对偶性(duality) 对偶规划问题的一个基本性质. 它表明了关于原问题存在最优解和对偶问题存在最优解之间的关系. 即:若原问题有最优解,则对偶问题也有最优解,且目标函数值相等. 反之亦成立.

松紧性(complementary slackness) 对偶规划问题的一个性质. 若 $\boldsymbol{x}^*$ 和 $\boldsymbol{y}^*$ 分别是原问题和对偶问题的可行解,则 $\boldsymbol{x}^*=(x_1^*,x_2^*,\cdots,x_n^*)^{\mathrm{T}}$ 和 $\boldsymbol{y}^*=(y_1^*,y_2^*,\cdots,y_m^*)^{\mathrm{T}}$ 分别是原问题和对偶问题的最优解的充分必要条件是:对于所有的 $i\in\{1,2,\cdots,m\}$ 有:

1. 若 $y_i^*>0$,则 $\sum_j a_{ij}x_j^*=b_i$.

2. 若 $\sum_j a_{ij}x_j^*<b_i$,则 $y_i^*=0$.

对偶线性规划问题解的基本性质(basic properties of dual linear programming solution) 互为对偶的两个线性规划问题之间的解的基本特征:

1. 两个问题同时有最优解,而且它们的目标函数值相等.

2. 两个问题同时没有可行解.

3. 一个问题有可行解,而无有限最优值,则另一个问题没有可行解.

对偶线性规划的非退化的基可行解(nondegenerate basic feasible solution of dual linear programming) 对偶线性规划的基本概念之一. 是使对偶线性规划问题的约束条件恰有 m 个等式成立的基可行解. 若 $\boldsymbol{y}^{(0)}$ 是对偶线性规划问题的一个基可行解,用 $\boldsymbol{y}^{(0)}$ 代入对偶线性规划问题中,取等号的式子恰好是 m 个. 称 $\boldsymbol{y}^{(0)}$ 是非退化的基可行解.

对偶线性规划的退化的基可行解(degeneracy in basic feasible solution of dual linear programming) 对偶线性规划的基本概念之一. 是使对偶线性规划问题的约束条件多于 m 个等式成立的基可行解. 若用对偶线性规划问题的一个基可行解 $\boldsymbol{y}^{(0)}$ 代入对偶线性规划问题中,取等号的式子多于 m 个,则称 $\boldsymbol{y}^{(0)}$ 是退化的基可行解. 若对偶线性规划问题有一组退化的基可行解,则称它为退化的对偶线性规划(问题). 在对偶线性规划退化时,对原线性规划问题施行对偶单纯形法(参见"对偶单纯形法")迭代,可能出现循环现象,这时可用布兰德规则(参见"布兰德规则")来避免循环.

退化的对偶线性规划(degeneracy in dual linear programming) 见"对偶线性规划的退化的基可行解".

对偶单纯形法(dual simplex method) 与单纯形法相类似的求解线性规划问题的迭代方法(参见"单纯形法"和"换基迭代"). 求解是在对偶规划问题的可行解集合的极点上进行迭代的. 设 B 为原标准线性规划问题的一个基,若 B 对应的原问题的检验数均小于或等于 0,则称 B 为原问题的一个正则基,B 所对应的基可行解为原问题的正则解. 对偶单纯形法的基本思路是:从原问题的一个正则解出发,在保持对偶问题的解始终是可行解的前提下,经过换基迭代,逐步使正则解变成原问题的可行解,从而得到原问题的最优解. 对偶单纯形法是勒姆克(Lemke, C. K.)于 1954 年提出的. 对偶单纯形法的计算步骤如下:

1. 引入松弛变量,把原问题化为标准型,列出初始单纯形表.

2. 如果检验数全部非正,检查 $\bar{\boldsymbol{b}}$ 列的各分量. 若 $\bar{\boldsymbol{b}}$ 列的各分量均非负,则得到原问题的最优解,计算终止;否则转入步骤 3.

3. 由 $\min\{i\,|\,b_{i0}<0\}=r$,确定对应的基变量 x_r 为出基变量. 同时,检查 x_r 所在行的各系数 $b_{rj}(j=1,2,\cdots,n)$. 若所有的 $b_{rj}\geqslant0$,则原问题无可行解,计

算终止；否则，由

$$s=\min\left\{t\mid\theta_t=\min\left\{\frac{\lambda_j}{b_{rj}}\middle|b_{rj}<0,1\leqslant j\leqslant n\right\}\right\}$$

确定 b_{rs}，相应的非基变量 x_s 为入基变量，同时确定 b_{rs}为主元，并记为 b_{rs}^*.

4. 以 b_{rs}为主元，进行初等行变换，换基迭代得到新的单纯形表，并返回步骤 2.

边际价格(marginal price)　亦称边际收益. 对偶线性规划问题中目标函数的增加值. 即在最优生产方案下，资源 b_i 增加一个单位资源时，目标函数 f 的增值：

$$y_i^*=\frac{\Delta f}{\Delta b_i}\quad(\text{这里 }\Delta b_i=1).$$

由弱对偶性说明，总产值不大于资源总的边际价格(参见"弱对偶性"). 由对偶性说明，当资源得到最优配置时，总产值等于总边际价格(参见"对偶性"). 由松紧性说明，如果资源 b_i 的边际价格大于零，是指在实现最优方案中资源 b_i 已被全部利用，资源 b_i 对最优方案来说是紧缺的(参见"松紧性"). 如果在实现最优方案时有

$$\sum_j a_{ij}x_j^*<b_i,$$

说明资源 b_i 未被全部利用，尚有剩余，因而资源 b_i 的边际价格为零. 总之，资源的边际价格依赖于决策(方案)，同种资源在不同的决策(方案)下，边际价格可能不同.

边际收益(marginal revenue)　即"边际价格".

影子价格(shadow price)　亦称最优计划价格或预测价格. 是资源在最优决策下边际价格的反映. 把某一经济结构中的某种资源在最优决策下的边际价格称为该资源的影子价格(参见"边际价格"). 设目标函数值的增量为 Δf^*，则

$$y_i^*=\frac{\Delta f^*}{\Delta b_i}.$$

若以微分近似代替差分，则

$$y_i^*=\frac{\partial f^*}{\partial b_i}.$$

当 $\Delta b_i=1$ 时，$y_i^*=\Delta f^*$，它表示资源 b_i 在最优决策下的边际价格. 因此，影子价格是资源在最优决策下边际价格的反映. 没有最优决策就没有影子价格. 影子价格只受经济结构本身客观条件所制约，不受人为因素的影响. 因此，资源的影子价格定量地反映了单位资源在最优生产方案中应为总收益提供的产值. 资源的影子价格也称为在最优方案中投入生产的资源的机会成本.

最优计划价格(optimal planning price)　即"影子价格".

预测价格(informing price)　即"影子价格".

机会成本(chance cost)　见"影子价格".

影子价格的求法(solution to shadow price)　求影子价格的方法. 一种利用线性规划原问题最终单纯形表 $T(B)$(参见"单纯形表")求得各资源的影子价格的方法. 在原问题最终单纯形表$T(B)$中，第 i 种资源(即第 i 个约束条件)对应的松弛变量 $x_{n+i}(i=1,2,\cdots,m)$的检验数的相反数就是第 i 种资源的影子价格. 影子价格的作用可分为以下几个方面：

1. 指出了企业挖潜革新、扬长避短的方向. 对于影子价格为正数的短线资源，一方面可在资源的影子价格大于该种资源的市场单价时，适量地增加购进这种资源，即可增加企业的总收益. 另一方面，如果能在生产工艺上革新，降低这种资源的消耗，将使企业增收节支，提高经济效益. 对于影子价格为零的长线资源，其剩余资源就是进一步发展生产的潜在优势，提供了增加生产以长线资源为主要资源的新产品的可能性.

2. 对紧缺资源实行"择优分配"的原则. 如各企业为了增加流动资金而向银行贷款，银行在发放贷款时，应首先考查各企业的资金边际利润率：

$$\text{资金边际利润率}=\frac{\text{总利润}-\text{投资量}}{\text{投资量}}.$$

设投资量为一个单位，则资金的边际利润率等于资金的影子价格减 1. 据此，银行可以择优扶持. 若资金的影子价格大于 1，则资金边际利润率为正，这对企业投资有经济效果，投资方案可行. 资金边际利润率越大，投资的效果就越显著. 若资金的影子价格为 1，则资金边际利润率为零，这对企业投资是不必要的. 若资金的影子价格小于 1，则资金边际利润率为负，这对企业投资是不可行的.

3. 可以预测产品价格. 买方要卖方的产品作为生产资源. 买方希望价格尽量的低，卖方希望价格尽量的高，双方都希望找到一个彼此都能接受的"合理价格". 该资源的影子价格就是买方的重要依据. 买方要求产品的价格必须小于该产品作为自己最优生产的资源的影子价格，否则将无利可图；卖方要求产品的价格必须大于自己生产的"成本"，否则要亏本. 因此，产品价格必在成本和影子价格之间. 可见，在经济管理中，如果能计算出某产品在不同经济结构中的影子价格，就可以从资源利用的角度为产品的价格预测出上限和下限.

4. 为决策者提供调整企业最优生产方案的信息. 在最终单纯形表中，第 j 种产品所对应的检验数为 $c_j-\boldsymbol{c}_BB^{-1}\boldsymbol{p}_j$，其中 c_j 表示第 j 种产品的单位价格，$\boldsymbol{p}_j$ 表示生产单位第 j 种产品所需要各种资源的数量，$\boldsymbol{c}_BB^{-1}\boldsymbol{p}_j$ 表示在最优生产方案下生产一个单位第 j 种产品所必须提供的产值. 由此，当 $c_j-\boldsymbol{c}_BB^{-1}\boldsymbol{p}_j<0$ 时，表示第 j 种产品的单位售价低于该产品必须提供的产值，此时生产第 j 种产品是不合

算的，不能生产该种产品；当 $c_j-\boldsymbol{c}_B B^{-1}\boldsymbol{p}_j\geqslant 0$ 时，表示第 j 种产品的单位售价高于该产品必须提供的产值，此时生产第 j 种产品将是有利可图的，该种产品应该投产.

5. 可以作为同类企业经济效益的评价标准之一.

灵敏度分析（sensitivity analysis） 考查线性规划问题的最优解对系数变化反应的灵敏程度. 它是在保证现行最优基不变的前提下，充分利用已得到的结果，确定线性规划数学模型中的系数 c_j,a_{ij},b_i 因受各种因素（政策变化、价格波动、运输能力、资源储备、工艺改进、设备更新等）影响而产生变化的允许范围，亦即考查系数变化是否影响基 B 的可行性和最优性. 灵敏度分析可分以下三个方面：

1. 目标函数中系数 c_j 发生变化，其他条件不变，确定 c_j 的允许变化范围：

1）若 c_j 为基变量的系数，则只要解不等式组 $\boldsymbol{c}_N-\bar{\boldsymbol{c}}_B B^{-1}N\leqslant\boldsymbol{0}$，便得 c_j 的允许变化范围 $[\alpha,\beta]$，其中 $\bar{\boldsymbol{c}}_B$ 为 c_j 变化后目标函数中基变量系数向量. 若 c_j 在 $[\alpha,\beta]$ 内变化，则 B 仍是最优基，最优解不变，而目标函数值变为 $\bar{\boldsymbol{c}}_B B^{-1}\boldsymbol{b}$；若 c_j 的变化超出了允许范围 $[\alpha,\beta]$，则 B 不再是最优基.

2）若 c_j 为非基变量的系数，则只要解不等式组 $\bar{\boldsymbol{c}}_j-\boldsymbol{c}_B B^{-1}\boldsymbol{p}_j\leqslant\boldsymbol{0}$，便得 c_j 的允许变化范围 $(0,\boldsymbol{c}_B B^{-1}\boldsymbol{p}_j]$，即 $\boldsymbol{c}_B B^{-1}\boldsymbol{p}_j$ 是 c_j 允许变化范围的上界，其中 $\bar{\boldsymbol{c}}_j$ 为 c_j 变化后的目标函数中非基变量系数向量. 若变化后 $\bar{\boldsymbol{c}}_j\leqslant\boldsymbol{c}_B B^{-1}\boldsymbol{p}_j$，则 B 仍是最优基，最优解不变，目标函数值也不变. 若 $\bar{\boldsymbol{c}}_j>\boldsymbol{c}_B B^{-1}\boldsymbol{p}_j$，则须改变现行方案.

3）若 c_j 分别为目标函数中基变量的系数和非基变量的系数，则只要解不等式组 $\bar{\boldsymbol{c}}_N-\bar{\boldsymbol{c}}_B B^{-1}N\leqslant\boldsymbol{0}$，便得到这些系数允许变化范围 $[\alpha',\beta']$，其中 $\bar{\boldsymbol{c}}_N$ 和 $\bar{\boldsymbol{c}}_B$ 分别为 c_j 变化后的目标函数中非基变量和基变量的系数向量.

2. 常数项 b_i 发生变化，其他条件不变，确定 b_i 的允许变化范围. 如果只是常数项 b_i 发生变化，并不影响 B 的正则性，而有可能影响 B 的可行性. 只要解 $B^{-1}\boldsymbol{b}\geqslant\boldsymbol{0}$，便可得到 b_i 的允许变化范围 $[\alpha,\beta]$. 若 b_i 在 $[\alpha,\beta]$ 中变化，则 B 仍是最优基，最优解不变，但 $\boldsymbol{b}$ 却随 b_i 的变化而改变，目标函数值也随之改变，分别为 $B^{-1}\bar{\boldsymbol{b}}$ 和 $\boldsymbol{c}_B B^{-1}\bar{\boldsymbol{b}}$，其中 $\bar{\boldsymbol{b}}$ 表示 b_i 变化后的 $\boldsymbol{b}$ 向量；若 b_i 的变化超出了允许变化范围 $[\alpha,\beta]$，则 B 不再是最优基，现行方案须改变.

3. 系数矩阵 A 中某元素 a_{ij} 发生变化，其他条件不变，确定 a_{ij} 的允许变化范围：

1）设非基变量的系数列 $\boldsymbol{p}_j$ 中的某元素 a_{rj} 变化为 $\bar{a}_{rj}$，不影响可行性，但最优性 $\boldsymbol{c}-\boldsymbol{c}_B B^{-1}A\leqslant\boldsymbol{0}$ 可能受到影响，只要解不等式组 $c_j-\boldsymbol{c}_B B^{-1}\bar{\boldsymbol{p}}_j\leqslant\boldsymbol{0}$，便得到 a_{rj} 的允许变化范围

$$\left[\frac{c_j-\sum\limits_{i\neq r}y_i a_{ij}}{y_r},+\infty\right],$$

其中 $\boldsymbol{y}$ 为单纯形表 $T(B)$ 中松弛变量所对应的检验数的相反数，且 $\boldsymbol{y}=\boldsymbol{c}_B B^{-1}=(y_1,y_2,\cdots,y_r,\cdots,y_m)$. 若 a_{rj} 在此范围内变化，则最优解不变，目标函数值也不变.

2）设基变量的系数列中某元素 a_{ij} 发生变化，这时，可行性和最优性都受到影响，应从单纯形表 $T(B)$ 中的具体变化来计算和分析.

3）添加新变量 x_t. 设目标函数中 x_t 的系数是 c_t，系数矩阵 A 中 x_t 对应列为 $\boldsymbol{P}_t$. 增加 x_t 不影响可行性，这时检验数行向量为 $(\boldsymbol{c}-\boldsymbol{c}_B B^{-1}A,c_t-\boldsymbol{c}_B B^{-1}\boldsymbol{P}_t)$. 因 B 是最优基，故 $\boldsymbol{c}-\boldsymbol{c}_B B^{-1}A\leqslant\boldsymbol{0}$. 所以，当 $c_t-\boldsymbol{c}_B B^{-1}\boldsymbol{p}_t\leqslant 0$ 时，B 仍是最优基，方案不变. 当 $c_t-\boldsymbol{c}_B B^{-1}\boldsymbol{p}_t>0$ 时，则 x_t 应成为基变量. 用单纯形法进行换基迭代，求出新的最优解.

4）添加新的约束条件. 一般地，这样做会增加基矩阵 B 的阶数，与原规划有很大差别. 可以在原规划最终单纯形表上求解新规划的最优解.

参数线性规划（parametric linear programming） 数学规划的一个分支. 它研究系数不是常数，而是在某范围内变化的参数的线性规划问题. 求解参数线性规划问题的目的就是求出参数在不同范围内对应的线性规划问题的最优解. 迈恩（Manne，A. S.）、萨蒂（Saaty，T. L.）和加斯（Gass，S.）依次在 1953 年、1954 年和 1955 年都研究过含参数的线性规划问题：

1. 目标函数 $f(\boldsymbol{x})=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ 中系数 $\boldsymbol{c}$ 含有参数的线性规划问题的一般形式是：

$$\begin{cases}\max\ \bar{\boldsymbol{c}}^{\mathrm{T}}x=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}+\delta\boldsymbol{c}^{*\mathrm{T}}\boldsymbol{x},\\ \text{s.t.}\ \ A\boldsymbol{x}=\boldsymbol{b},\\ \qquad \boldsymbol{x}\geqslant\boldsymbol{0}.\end{cases}$$

对应基 B 的单纯形表 $T(B)$ 为

$c_j\rightarrow$			$c_1+\delta c_1^*$	$c_2+\delta c_2^*$	$\cdots$	$c_n+\delta c_n^*$
$\bar{\boldsymbol{c}}_B$	$\boldsymbol{x}_B$	$\bar{\boldsymbol{b}}$	x_1	x_2	$\cdots$	x_n
	x_{j1}	b_{10}	b_{11}	b_{12}	$\cdots$	b_{1n}
	x_{j2}	b_{20}	b_{21}	b_{22}	$\cdots$	b_{2n}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	x_{jm}	b_{m0}	b_{m1}	b_{m2}	$\cdots$	b_{mn}
	$-f$	$-b_0-b_0^*\delta$	$\lambda_1+\lambda_1^*\delta$	$\lambda_2+\lambda_2^*\delta$	$\cdots$	$\lambda_n+\lambda_n^*\delta$

当 $\underline{\delta}_B\leqslant\delta\leqslant\bar{\delta}_B$ 时，$T(B)$ 是最优基 B 对应的单纯形表. $\underline{\delta}_B$ 和 $\bar{\delta}_B$ 分别称为基 B 的下特征数和上特征数，区间 $[\underline{\delta}_B,\bar{\delta}_B]$ 称为 B 的最优区间. 其中

$$\underline{\delta}_B=\begin{cases}\max\left\{-\frac{\lambda_j}{\lambda_j^*}\middle|\lambda_j^*<0\right\}, & 若存在\ \lambda_j^*<0,\\ -\infty, & 若不存在\ \lambda_j^*<0,\end{cases}$$

$$\bar{\delta}_B=\begin{cases}\min\left\{-\frac{\lambda_j}{\lambda_j^*}\middle|\lambda_j^*>0\right\}, & 若存在\ \lambda_j^*>0,\\ +\infty, & 若不存在\ \lambda_j^*>0.\end{cases}$$

若 $\delta>\bar{\delta}_B$，则存在 $\lambda_j+\lambda_j^*\delta>0\ (1\leqslant j\leqslant n)$，因此基 B 不是最优基. 这时有两种情形：

1）检验数 $\lambda_j+\lambda_j^*\delta>0$ 对应的系数列

$$\begin{pmatrix}b_{1j}\\b_{2j}\\\vdots\\b_{mj}\end{pmatrix}\leqslant 0,$$

由单纯形表知，参数线性规划问题无有限最优解.

2）检验数 $\lambda_j+\lambda_j^*\delta>0$ 对应的系数列

$$\begin{pmatrix}b_{1j}\\b_{2j}\\\vdots\\b_{mj}\end{pmatrix}\nleqq 0,$$

为求 $\delta>\bar{\delta}_B$ 时的最优基，x_j 应作为入基变量. 用单纯形法进行换基迭代，得到新的最优基对应的单纯形表，进一步得到最优区间. 如果 $\delta<\underline{\delta}_B$，重复上述讨论.

2. 约束条件常数项 b 含有参数的线性规划问题的一般形式是

$$\begin{cases}\max\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},\\ \text{s.t.}\ A\boldsymbol{x}=\bar{\boldsymbol{b}},\\ \qquad \boldsymbol{x}\geqslant\boldsymbol{0},\end{cases}$$

其中 $\bar{\boldsymbol{b}}=(\boldsymbol{b}+\theta\boldsymbol{b}^*)=(b_1+b_1^*\theta,b_2+b_2^*\theta,\cdots,b_m+b_m^*\theta)^{\mathrm{T}}$. 对应基 B 的单纯形表 $T(B)$ 为：

$c_j\to$			c_1	c_2	$\cdots$	c_n	
c_B	x_B	$\bar{b}$	x_1	x_2	$\cdots$	x_n	
	x_{j1}	$b_{10}+b_{10}^*\theta$	b_{11}	b_{12}	$\cdots$	b_{1n}	
	x_{j2}	$b_{20}+b_{20}^*\theta$	b_{21}	b_{22}	$\cdots$	b_{2n}	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
	x_{jm}	$b_{m0}+b_{m0}^*\theta$	b_{m1}	b_{m2}	$\cdots$	b_{mn}	
	$-f$	$-f_0-f_0^*\theta$	λ_1	λ_2	$\cdots$	λ_n	

当 $\underline{\theta}_B\leqslant\theta\leqslant\bar{\theta}_B$ 时，$T(B)$ 是最优基 B 的单纯形表. $\underline{\theta}_B$ 和 $\bar{\theta}_B$ 分别称为关于 θ 的下特征数和上特征数，$\underline{\theta}_B,\bar{\theta}_B$ 构成的区间为最优区间，其中

$$\underline{\theta}_B=\begin{cases}\max\left\{-\frac{b_{i0}}{b_{i0}^*}\middle|b_{i0}^*>0\right\}, & 若存在\ b_{i0}^*>0,\\ -\infty, & 若不存在\ b_{i0}^*>0,\end{cases}$$

$$\bar{\theta}_B=\begin{cases}\min\left\{-\frac{b_{i0}}{b_{i0}^*}\middle|b_{i0}^*<0\right\}, & 若存在\ b_{i0}^*<0,\\ +\infty, & 若不存在\ b_{i0}^*<0.\end{cases}$$

若 $\theta>\bar{\theta}_B$，则必存在 $b_{i0}+b_{i0}^*\theta<0\ (1\leqslant i\leqslant m)$. 此时有两种情形：

1）$b_{i0}+b_{i0}^*\theta<0$，对应的系数行向量 $(b_{i1},b_{i2},\cdots,b_{in})\geqslant 0$，由对偶单纯形法知，$\theta>\bar{\theta}_B$ 时，参数线性规划问题无可行解.

2）$b_{i0}+b_{i0}^*\theta<0$，对应的系数行向量 $(b_{i1},b_{i2},\cdots,b_{in})\ngeqq 0$，由对偶单纯形法知，$x_i$ 为出基变量，用对偶单纯形法进行换基迭代，可求得新的单纯形表，进一步得到新的最优区间. 如果 $\theta<\underline{\theta}_B$，重复上述讨论.

基的下特征数（low characteristic number of base）　见“参数线性规划”.

基的上特征数（above characteristic number of base）　见“参数线性规划”.

关于 θ 的下特征数（low characteristic number on θ）　见“参数线性规划”.

关于 θ 的上特征数（above characteristic number on θ）　见“参数线性规划”.

特殊的线性规划问题

特殊的线性规划问题（particular linear programming problem）　一类重要而实用的线性规划问题. 指运输问题、指派问题以及变量有上界的线性规划问题等. 单纯形法虽然适用于求解任何线性规划问题，但这些特殊的线性规划问题有比单纯形法更简便、更有效的求解方法.

运输问题（transportation problem）　特殊的线性规划问题. 分为产销平衡运输问题和产销不平衡运输问题. 它的数学模型是为了解决调运物资而建立起来的，除了用单纯形法求解之外，还有比单纯形法更简便、更有效的解法. 运输问题的研究，出现在线性规划的一般理论和单纯形法之前，希契科克（Hitchcock，F. L.）、坎托罗维奇（Канторович，Л. В.）和库普曼斯（Koopmans，T. C.）依次于 1941 年、1942 年和 1947 年先后对此进行了研究，到 1951 年才由丹齐克（Dantzig，G. B.）把单纯形法用来解决运输问题.

产销平衡运输问题（balanced transportation problem）　总产量与总销量相等的运输问题. 设将某种物资从 m 个产地 $A_1,A_2,\cdots,A_m$ 运送到 n 个销地 $B_1,B_2,\cdots,B_n$，其中产量分别为 $a_1,a_2,\cdots,a_m$，销量分别为 $b_1,b_2,\cdots,b_n$，并且

$$\sum_{i=1}^{m}a_i=\sum_{j=1}^{n}b_j.$$

已知从第 i 个产地到第 j 个销地的每吨公里运费为 $c_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$，产销平衡运输问题就是要解决如何调运才能既满足上述需要，又使总的运输吨公里费用达到最少. 上面这些给定的数据

通常用产销平衡表和单位运价表表示：

产销平衡表 **单位运价表**

产＼销 地	B_1 B_2 ⋯ B_n	产量/吨	B_1 B_2 ⋯ B_n
A_1		a_1	c_{11} c_{12} ⋯ c_{1n}
A_2		a_2	c_{21} c_{22} ⋯ c_{2n}
⋮		⋮	⋮ ⋮ ⋱ ⋮
A_m		a_m	c_{m1} c_{m2} ⋯ c_{mn}
销量/吨	b_1 b_2 ⋯ b_n		

产销平衡运输问题的数学模型为

$$\begin{cases}\min \sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij},\\ \text{s.t. } \sum_{j=1}^{n}x_{ij}=a_i \quad (i=1,2,\cdots,m),\\ \sum_{i=1}^{m}x_{ij}=b_j \quad (j=1,2,\cdots,n),\\ x_{ij}\geqslant 0 \quad \begin{pmatrix} i=1,2,\cdots,m\\ j=1,2,\cdots,n\end{pmatrix},\end{cases}$$

其中$\sum_{i=1}^{m}a_i=\sum_{j=1}^{n}b_j$. 或用矩阵形式表示为

$$\begin{cases}\min\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},\\ \text{s.t. } A\boldsymbol{x}=\boldsymbol{b},\\ \boldsymbol{x}\geqslant\boldsymbol{0},\end{cases}$$

其中 $\boldsymbol{c}=(c_{11},c_{12},\cdots,c_{1n},c_{21},c_{22},\cdots,c_{2n},\cdots,c_{m1},c_{m2},\cdots,c_{mn})^{\mathrm{T}}$,

$$A=\begin{pmatrix}1\ 1\cdots 1 & & & \\ & 1\ 1\cdots 1 & & \\ & & \ddots & \\ & & & 1\ 1\cdots 1\\ 1 & 1 & & 1\\ \ 1 & \ 1 & & \ 1\\ \ \ddots & \ \ddots & \cdots & \ \ddots\\ \ \ 1 & \ \ 1 & & \ \ 1\end{pmatrix}\begin{matrix}\left.\right\} m\text{ 行}\\ \left.\right\} n\text{ 行}\end{matrix}$$

$=(\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_{mn})$,

$\boldsymbol{b}=(a_1,\cdots,a_m,b_1,\cdots,b_n)^{\mathrm{T}}$,

$\boldsymbol{x}=(x_{11},x_{12},\cdots,x_{1n},x_{21},\cdots,x_{mn})^{\mathrm{T}}$.

产销不平衡运输问题(unbalanced transportation problem)　总产量与总销量不相等的运输问题. 即

$$\sum_{i=1}^{m}a_i\neq\sum_{j=1}^{n}b_j$$

的运输问题(参见“产销平衡运输问题”). 当产大于销时，只要增加一个假想的销地 B_{n+1}(实际上是贮存)，且该销地总需要量为

$$b_{n+1}=\sum_{i=1}^{m}a_i-\sum_{j=1}^{n}b_j,$$

而在单位运价表中，从各产地到假想销地的单位运价为 $c_{i,n+1}=0$，这种产销不平衡的运输问题就转化成一个产销平衡的运输问题. 当销大于产时，可以在产销平衡表中增加一个假想的产地 A_{m+1}，该地产量为

$$a_{m+1}=\sum_{j=1}^{n}b_j-\sum_{i=1}^{m}a_i,$$

在单位运价表上，令从该假想产地到各销地的运价 $c_{m+1,j}=0$，同样将产销不平衡运输问题转化为一个产销平衡的运输问题. 当产大于销时，产销不平衡运输问题的数学表达式为

$$\begin{cases}\min \sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij},\\ \text{s.t. } \sum_{j=1}^{n}x_{ij}\leqslant a_i \quad (i=1,2,\cdots,m),\\ \sum_{i=1}^{m}x_{ij}=b_j \quad (j=1,2,\cdots,n),\\ x_{ij}\geqslant 0 \quad \begin{pmatrix} i=1,2,\cdots,m\\ j=1,2,\cdots,n\end{pmatrix},\end{cases}$$

其中

$$\sum_{i=1}^{m}a_i\geqslant\sum_{j=1}^{n}b_j.$$

当销大于产时，其数学表达式为

$$\begin{cases}\min \sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij},\\ \text{s.t. } \sum_{j=1}^{n}x_{ij}=a_i \quad (i=1,2,\cdots,m),\\ \sum_{i=1}^{m}x_{ij}\leqslant b_j \quad (j=1,2,\cdots,n),\\ x_{ij}\geqslant 0 \quad \begin{pmatrix} i=1,2,\cdots,m\\ j=1,2,\cdots,n\end{pmatrix},\end{cases}$$

其中

$$\sum_{i=1}^{m}a_i\leqslant\sum_{j=1}^{n}b_j.$$

运输模型的对偶(dual of transportation problem)　运输问题的线性规划模型的对偶模型. 它的数学表达式为

$$\begin{cases}\max \sum_{i=1}^{m}a_iu_i+\sum_{j=1}^{n}b_jv_j,\\ \text{s.t. } u_i+v_j\leqslant c_{ij},\\ u_i,v_j\text{ 无约束}\begin{pmatrix} i=1,2,\cdots,m\\ j=1,2,\cdots,n\end{pmatrix},\end{cases}$$

其中 u_i,v_j 称为对偶变量.

运输问题的对偶变量(dual variable of transportation problem)　见“运输模型的对偶”.

运输问题基的特征(features of the transportation problem basis)　运输问题的一个基本特性. 运输问题中任何一个基都对应系数矩阵的 $m+n-1$

个线性无关的列向量，从而任何一个基可行解都有 $m+n-1$ 个基变量.

表作业法(hitchock method) 一种与单纯形法相类似的求解运输问题的方法. 在表上先确定一个初始方案，然后反复进行调整，最后得到最优解. 表的构成参见"产销平衡运输问题"中的产销平衡表和单位运价表. 表作业法的步骤如下：

1. 用最小元素法制定初始方案(参见"最小元素法").

2. 求出检验数，判别方案是否最优. 求检验数的方法有闭回路法(参见"闭回路法")、位势法(参见"位势法")和加圈法(参见"加圈法").

3. 求出调整量，在闭回路上进行方案的调整(参见"闭回路法").

表作业法的换基迭代，是在调运表上负检验数对应的空格所在的闭回路上进行的(参见"换基迭代"). 调整后，空格对应的非基变量值由零增到 θ，成为新基可行解的基变量，而原方案中这条闭回路的第奇数次拐角点所对应的基变量值中有一个为零，改为空格，成为新基可行解中的非基变量. 如果同时出现几个零，规定只将其中最上方那行的最左边出现的那个零改为空格，其他的零均要填上，仍以基变量对待. 这就保证了其中填数的格子(即基变量)仍为 $m+n-1$ 个. 新方案相应的总运费下降数值为对应于空格的检验数与调整数之积的绝对值. 继续对新方案判别、调整，经过有限次直至所有的检验数都非负，便得出最优解. 用表作业法解运输问题，可能有多个最优解.

最小元素法(minimum element method) 一种求解运输问题初始调运方案的方法. 该方法是按照运费最小优先供应原则来确定初始调运方案. 具体的作法是：从运价表中观察最小的运价，在未被划去的运价中找最小运价，直到产量全部运完，需求量完全满足为止. 在表作业过程中，平衡表上每填一个数字，运价表上就划去一行或一列. 产量大于销量时，运价表上划去列，销量大于产量时，则划去行. 但由于最后一个数字要求平衡，即产量与销量相等，所以在调运表上填完最后一个数字时，在运价表上同时划去一行和一列，因而在调运表上共须填 $m+n-1$ 个数字，即 $m+n-1$ 个基变量. 需要注意的是：有时在作业的中间过程中，产量与销量相等，这时只能划去一行或一列，不能同时都划去，另一行或另一列待以后在调运表中填写数字时再划去，且在调运表上填写数字"0". 这个零是基变量，取 0 值，这个格子要看成有数字的格子，不是空格，它与非基变量的空格不同. 否则，会造成调运表中数字少于 $m+n-1$ 个. 总之，任何一个平衡运输问题都可以用最小元素法求出一个初始可行解. 又因运输问题的目标函数有下界(且不为负)，所以平衡运输问题必有最优解. 然而用最小元素法求得的初始方案不一定就是最优方案.

最大元素法(maximum element method) 一种求解运输问题初始方案的方法. 该方法是按最大利润优先供应的原则，制定初始调运方案. 具体作法与最小元素法的具体作法相类似(参见"最小元素法").

西北角法(northwest corner method) 一种求解运输问题初始方案的方法. 方法是从平衡表的西北角开始确定初始方案. 即在方格 (A_1,B_1) 中填入 (b_1,a_1) 的较小值. 有两种可能：

1. 若 $b_1<a_1$，则把 (b_2,a_1-b_1) 的较小数填入 (A_1,B_2).

2. 若 $b_1>a_1$，则把 (b_1-a_1,a_2) 的较小数填入 (A_2,B_1).

如此继续，直到表的东南角，把各地的产量全部运完为止，从而确定出初始调运方案.

闭回路(loop) 求解运输问题时产生的闭合回路. 在调运方案表中，从一个空格出发，沿水平或垂直方向前进；遇到一个适当的有数字的格子，则转 90°前进；这样必会又遇到一个适当的有数字的格子，同样再转 90°向前进. 经若干次后，必然会回到出发的那个空格. 这样形成一条由水平线与垂直线构成的封闭折线，称为一个闭回路.

空格的检验数(test number of cell) 在运输问题中，用以检验调运方案最优与否的数字. 它等于过空格的闭回路中第偶数次拐角点运价的总和减去第奇数次拐角点运价的总和.

闭回路法(loop method) 在运输问题中，用来求非基变量的检验数和调整调运方案的方法. 一般地，在调运方案中作一个闭回路(参见"闭回路")，计算空格的检验数(参见"空格的检验数")，对调运方案进行调整. 若所有的检验数都非负，则该调运方案为最优方案；否则就需要继续调整，直至所有检验数均非负.

位势法(potential method) 在运输问题中，用来求非基变量的检验数和调整调运方案的方法. 一般地，在调运方案中，凡有数字的格，都换上单位运价表中对应的运价，并在表的最右面和最下面增加一行和一列，通常用 $u_i(i=1,2,\cdots,m)$ 和 $v_j(j=1,2,\cdots,n)$ 表示，分别称为第 i 行和第 j 列的位势. 若第 i 行第 j 列为数字格，则 u_i 和 v_j 之和等于表中对应的运价数，即 $u_i+v_j=c_{ij}$. 对已知的初始方案，只要先给出一个位势，则其他行、其他列的位势存在而且惟一. 由 $u_i+v_j=c_{ij}$，即可求出 u_i 和 v_j. 任一空格的检验数为 $\lambda_{ij}=c_{ij}-(u_i+v_j)$. 若所有的检验数都非负，则该调运方案为最优方案，否则就需要调整，直至所

有检验数均非负为止. 调整的方法与闭回路法相同.

位势(potential)　见“位势法”.

加圈法(adding loop method)　在运输问题中，一次性求出全部非基变量的检验数和调整调运方案的方法. 一般地，在运价表中，把相应于有调运数字的费用加上圈，再在此表上，对同行(或同列)中的各数减去(或加上)同一个数，反复进行，使加圈的数字全部为零. 此时，表中未加圈的其他各数便是全部的检验数. 若所有的检验数都非负，则该调运方案为最优方案；否则需要调整，直至所有检验数均非负.

表作业法的换基迭代(changing basis iteration of operation method on tableau)　一种求解运输问题过程中的迭代技巧. 指在调运表上对负检验数相应的空格所在的闭回路上进行的迭代. 先把第一个出现负检验数的空格所对应的非基变量作为入基变量，并使这些非基变量的值由零增到调整量. 为了保持平衡，在这空格的闭回路上取调整量 $\theta=\min\{$第奇数次拐角点的调运量$\}$，然后对闭回路中的每个第奇数次拐角点的调运量各减去调整量 θ，对每个第偶数次拐角点的调运量加上调整量 θ，闭回路外的调运量均不变. 这样，便得到一个新的调运方案.

图作业法(graphical method for transportation)　一种求解运输问题的方法. 是在一个标有产地、销地，以及连结产地和销地的网络交通图进行物资调运的方法. 它是一种迭代法，即先找第一个流向图(初始方案)，进行反复调整，直至得到最优流向图(最优方案)为止. 图作业法是中国运输工作者在实际工作中创造的一种解决运输问题的简便易行且利于推广的方法. 图作业法的步骤如下：

1. 求第一个流向图(初始方案). 首先注意避免对流，如果交通图上没有圈，不可能有迂回出现，依据就近调拨的原则，即“端点始，供需归邻站，各点供需为零止”，便可做出无对流的第一个流向图. 如果交通图上有圈，利用破圈法(或避圈法)，便可得到一个没有对流的流向图.

2. 判别第一个流向图是否最优. 其方法是：计算这个流向图的每一个仅含一条无流向边的圈上内外圈流向的总长. 若总长都不超过这个圈的一半，则这个流向图就是最优的. 否则，则转入步骤 3.

3. 对现有的流向图进行调整. 方法是：若内(外)圈之长大于整个圈长的一半，则内(外)圈各流量都减去内(外)圈中最小的流量，所有外(内)圈和空圈都按外(内)圈方向加上这个最小流量.

4. 重复步骤 2 和步骤 3，直至得到最优流向图为止.

流向(flow direction)　运输问题的基本概念之一. 指在运输问题中，从产地到销地的调运方向. 在交通图上，用“○”表示产地，○中的数字为该产地的产量，用“□”表示销地，□中的数字为该销地的销量. 从某产地到某销地的调运方向用“→”表示，称为流向，画在前进方向的右侧.

流量(flux)　运输问题的基本概念之一. 指在运输问题中，流向的运输量. 在交通图上记在“→”的右边.

流向图(flow diagram)　运输问题的基本概念之一. 指在运输问题中，标明产地的产量、销地的销量，以及流向和流量的交通图.

对流(convection)　运输问题的基本概念之一. 指在运输问题中，在一段线路上有同一种物资相对运输.

迂回(detour)　运输问题的基本概念之一. 指在运输问题中，在有圈的交通图中走了大半圈. 若在某有圈流向图中给出了一个调运方案，使得在某个圈中同一流向(顺时针或逆时针)的边长之和大于整个圈长的一半，则称这种现象为迂回.

转运问题(transshipment problem)　一类运输问题. 求一个运输方案，使得产地的货物都能运到销地去，并且使总的运输吨公里费用达到最小的问题. 设已知一个连通的无向图 $G=[V,E]$，图上的每一个顶点 v_i 对应一个实数 a_i($a_i>0$ 时，v_i 为产地；$a_i<0$ 时，v_i 为销地；$a_i=0$ 时，v_i 为中转点)，且满足

$$\sum_{i=1}^{n} a_i = 0.$$

另外，每一条边 $e_j=[s,t]\in E$ 有一个长度 c_{st}. 把每一条边 $l_j=[s,t]$看成两条弧(s,t)与(t,s)，它们的长度都等于 c_{st}. 转运问题的数学模型为

$$\begin{cases}\min \sum\limits_{(s,t)\in E} c_{st}x_{st}, \\ \text{s.t.} \sum\limits_{(i,j)\in E} x_{ij} - \sum\limits_{(k,i)\in E} x_{ki} = a_i \\ \qquad (i=1,2,\cdots,m), \\ \qquad x_{ij}\geqslant 0 \ ((i,j)\in E),\end{cases}$$

其中 x_{st}表示通过弧(s,t)的货物的数量，$\sum\limits_{(i,j)\in E} x_{ij}$ 表示从 v_i 运出的货物的总数，$\sum\limits_{(k,i)\in E} x_{ki}$ 表示运到 v_i 的货物的总数，而

$$\sum_{(i,j)\in E} x_{ij} - \sum_{(k,i)\in E} x_{ki}$$

就是 v_i 的“纯运出量”.

指派问题(assignment problem)　亦称分派问题或分配问题. 运输问题的特殊情形，也是 0-1 规划的特殊情形(参见“0-1 规划”). 一般有：

1. 设有 n 项加工任务，需要分别安排在 m 台机器上进行加工($m>n$)，由于各机器的生产效率不同，应如何安排，以使总的加工时间最少(或收益最大).

2. 把 m 个推销员分派到 n 个地方去推销产品

($m=n$),由于每人对不同地方的工作适应能力不同,如何分派,以使总效率最高.

3. 某工程有 m 个招标项目,现有 n 个公司进行投标($m<n$),如果每个项目只能由一个公司担任,应如何选取投标的公司,使总费用最少(或总收益最大).

分派问题(dispatch problem)　即"指派问题".

分配问题(distribution problem)　即"指派问题".

指派问题的数学模型(mathematical model of assignment problem)　运输问题数学模型的特殊形式.在运输问题中,当 $a_i=b_i=1$ 时,则它的数学模型就是指派问题的数学模型.其一般形式为:求一组 $x_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$,使

$$\begin{cases}\min \ \sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij},\\ \text{s.t.}\ \ \sum_{j=1}^{n}x_{ij}=1\quad(i=1,2,\cdots,m),\\ \qquad \sum_{i=1}^{m}x_{ij}=1\quad(j=1,2,\cdots,n),\\ \qquad x_{ij}\text{ 为 0 或 1 }\begin{pmatrix}i=1,2,\cdots,m\\ j=1,2,\cdots,n\end{pmatrix},\end{cases}$$

其中 $c_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$表示时间、费用、距离等,且 $c_{ij}\geqslant 0$;x_{ij}表示决策变量:

$$x_{ij}=\begin{cases}1,\ \text{当指派第 } i \text{ 人担任第 } j \text{ 项工作},\\ 0,\ \text{当不指派第 } i \text{ 人担任第 } j \text{ 项工作}.\end{cases}$$

效率矩阵(efficiency matrix)　亦称价值矩阵.指派问题中的重要参数.是由指派问题的数学模型中元素 $c_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$构成的矩阵

$$(c_{ij})_{m\times n}=\begin{pmatrix}c_{11} & c_{12} & \cdots & c_{1n}\\ c_{21} & c_{22} & \cdots & c_{2n}\\ \vdots & \vdots & & \vdots\\ c_{m1} & c_{m2} & \cdots & c_{mn}\end{pmatrix}.$$

价值矩阵(value matrix)　即"效率矩阵".

最优指派(optimal assignment)　指派问题中,使目标函数取得最优值(最大值或最小值)的指派.即在指派问题的一个可行解矩阵(x_{ij})中,各行各列都只有一个元素是1,其余都是0(其中1表示指派,0表示不指派)的问题的解所对应的指派.这个可行解 x_{ij}使目标函数

$$f=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}$$

取最大值(或最小值——视问题中目标函数是求最大或求最小而定).最优指派的最优解有一个很重要的性质:如果从效率矩阵(c_{ij})任何一行(列)的各元素中减去该行(列)的最小元素(或加上某一正数),所得到的新效率矩阵(b_{ij})所对应的指派问题的最优解与原问题的最优解相同.该性质是匈牙利法的理论依据(参见"匈牙利法").指派问题的最优解可能不惟一.

匈牙利法(Hungarian method)　求解指派问题的有效算法之一.它首先由匈牙利数学家柯尼希(König,D.)提出.其具体方法步骤是:

1. 变换效率矩阵(c_{ij}),使变换后的效率矩阵(b_{ij})的每行每列至少有一个零元素.即对(c_{ij})的每行和每列,分别减去该行和该列的最小元素.

2. 求最优解.在(b_{ij})上,从零元素最少的行开始,对一个零元素标注 0^*,表示一种指派.同时,对同行同列的其他零元素标注 φ,以防止下一个指派落在此行此列上.若该行被考虑的零元素(已标注 φ 的零元素不再考虑)多于一个,则标注 0^* 的零元素所在之列应当是零元素最少的.反复进行,直至(b_{ij})上所有的零元素都已标注为止.若最后能得到 n 个 0^*,则必然是不同行不同列上的 n 个零元素.将这 n 个 0^* 改为1,其他元素改为零,就形成一个完整的指派方案,得到指派问题的一个最优解,计算终止.否则,则转入步骤3.

3. 作能覆盖所有零元素(标注 φ 和 0^*)的最少直线集合.对没有 0^* 的行打√号,对打√号行上所有有零元素的列打√号,再对打√号列上有 0^* 的行打√号,直到得不出新的打√号的行列为止.对没有打√号的行画横线、所有打√号的列画竖线,其直线的条数等于(b_{ij})中 0^* 的个数.

4. 在未被直线覆盖的元素中找出最小元素,并对没画直线行的各元素减去这最小元素,对画直线列的各元素都加上这最小元素,变换成新的效率矩阵,并返回步骤2,重复进行,直至求得最优解为止.

如果对指派问题数学模型中的目标函数

$$f=\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij}$$

是求最大值,可作如下的变换:设构造一个新的效率矩阵($M-c_{ij}$),并取 $M=\max\{c_{ij}\}$,这时新矩阵的元素 $M-c_{ij}\geqslant 0$,则

$$\begin{aligned}\max\ f&=\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij}\\ &=nM-\min\sum_{i=1}^{n}\sum_{j=1}^{n}(M-c_{ij})x_{ij}.\end{aligned}$$

如果效率矩阵的 $m\neq n$,可先增加虚设的零行(列),使效率矩阵变成方阵,即当 $m<n$ 时为

$$\begin{pmatrix}c_{11} & c_{12} & \cdots & c_{1n}\\ \vdots & \vdots & & \vdots\\ c_{m1} & c_{m2} & \cdots & c_{mn}\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 0\end{pmatrix}\begin{matrix}\\ \\ \\ \left.\begin{matrix}\\ \\ \\ \end{matrix}\right\}(n-m)\text{行}\end{matrix},$$

当 $m>n$ 时为

$$\underbrace{\begin{pmatrix} c_{11} & \cdots & c_{1n} & 0 & \cdots & 0 \\ c_{21} & \cdots & c_{2n} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ c_{m1} & \cdots & c_{mn} & 0 & \cdots & 0 \end{pmatrix}}_{(m-n)\text{列}},$$

然后再利用匈牙利法求解.

变量有上界的线性规划问题(linear programming problem of upper bounded variable) 决策变量具有上界的线性规划问题. 设一般地,线性规划问题是

$$\begin{cases} \max\limits_{(\min)} \ \sum\limits_{j=1}^{n} c_j x_j, \\ \text{s.t.} \ \sum\limits_{j=1}^{n} a_{ij} x_j = b_i \quad (i=1,2,\cdots,m), \\ \qquad r_j \leqslant x_j \leqslant d_j \quad (j=1,2,\cdots,n), \end{cases}$$

则变量有上界的线性规划问题是

$$\begin{cases} \max\limits_{(\min)} \ c + \sum\limits_{j=1}^{n} c_j y_j, \\ \text{s.t.} \ \sum\limits_{j=1}^{n} a_{ij} y_j = f_i \quad (i=1,2,\cdots,m), \\ \qquad 0 \leqslant y_j \leqslant q_j \quad (j=1,2,\cdots,n), \end{cases}$$

其中

$$y_j = x_j - r_j, \qquad c = \sum_{j=1}^{n} c_j r_j,$$

$$f_i = b_i - \sum_{j=1}^{n} a_{ij} r_j, \quad q_j = d_j - r_j.$$

变量有上界的线性规划问题是丹齐克(Dantzig,G. B.)在研究工作任务分配问题时提出来的,并且给出一个解法. 这个解法实际上还是单纯形法,只是在迭代时作了一些改变. 同样,对对偶单纯形法作一些类似的改变,也能用来解变量有上界的线性规划问题,而且这些方法也能用来求解变量有上界的运输问题.

变量有上界的线性规划问题的解法(solution to linear programming problem of upper bounded variable) 求解变量有上界的线性规划问题

$$\begin{cases} \max \ \sum\limits_{j=1}^{n} c_j x_j, \\ \text{s.t.} \ \sum\limits_{j=1}^{n} a_{ij} x_j = b_i \quad (i=1,2,\cdots,m), \\ \qquad 0 \leqslant x_j \leqslant q_j \quad (j=1,2,\cdots,n) \end{cases}$$

的方法. 基本思路是:首先不考虑上界的约束,用单纯形法求解(参见"单纯形法"和"换基迭代"). 若得到的最优解满足 $0\leqslant x_j\leqslant q_j$,则这个最优解就是变量有上界约束的线性规划问题的最优解,计算终止. 否则,在换基迭代时,既要保证满足入基变量的非负约束和上界约束,还要保证满足入基后可能受到影响的那些基变量非负约束和上界约束,如此,有选择地确定出基变量. 反复进行,直至求得最优解或判定无有限最优解为止. 换基迭代步骤为:

1. 按单纯形法确定入基变量后,对入基变量 x_s 计算 $\min\{q_s,\theta,\beta\}$,其中 q_s 是 x_s 的上界. $\bar{\boldsymbol{p}}_s$ 列中若存在正分量,则取

$$\theta = \min\left\{\frac{b_{i0}}{b_{is}} \middle| b_{is} > 0\right\};$$

若不存在正分量,则取 $\theta=+\infty$. $\bar{\boldsymbol{p}}_s$ 列中若存在负分量,则取

$$\beta = \min\left\{\frac{q_s - b_{i0}}{-b_{is}} \middle| b_{is} < 0\right\};$$

若不存在负分量,则取 $\beta=+\infty$. 这样,就能满足可行性.

2. 换基准则. 若最小值为 q_s,则以 $x_s=q_s-y_s$ 代入,更新单纯形表,不需换基迭代. 若最小值为 θ,则按单纯形法进行换基迭代. 若最小值为 β,并设 β 出现于基变量 x_k 所在的行,则先以 $x_k=q_k-y_k$ 代入,更新单纯形表后,以 y_k 代替 x_k 为出基变量,再按单纯形法进行换基迭代. 若最小值为 $+\infty$,则问题无有限最优解.

大规模线性规划问题(large scale linear programming problem) 亦称大型线性规划问题. 变量个数和约束条件的个数都很大的线性规划问题. 求解这类线性规划问题可用:

1. 直接法(参见"一般稀疏矩阵法").

2. 分解算法(参见"丹齐克-沃尔夫分解算法").

大型线性规划问题(large scale linear programming problem) 即"大规模线性规划问题".

一般稀疏矩阵法(general sparse matrix method) 求解大规模线性规划问题的一种直接方法. 许多规模较大的线性规划问题的约束条件的系数矩阵是高度稀疏的,利用这一特点,运用矩阵代数的一些技巧(如矩阵的分解),结合单纯形法来求解.

可分解的线性规划问题(decomposed linear programming problem) 大规模线性规划问题中的一类特殊形式的问题. 它的一种形式是:求 $x_1,x_2,\cdots,x_n$,满足约束条件

$$\begin{cases} B_1\boldsymbol{x}_1 & & & =\boldsymbol{b}_1, \\ & B_2\boldsymbol{x}_2 & & =\boldsymbol{b}_2, \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ & & B_n\boldsymbol{x}_n & =\boldsymbol{b}_n, \\ A_1\boldsymbol{x}_1 + A_2\boldsymbol{x}_2 + \cdots + A_n\boldsymbol{x}_n = \boldsymbol{b}, \\ \boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n \geqslant \boldsymbol{0}, \end{cases}$$

并且使 $f=\boldsymbol{c}_1\boldsymbol{x}_1+\boldsymbol{c}_2\boldsymbol{x}_2+\cdots+\boldsymbol{c}_n\boldsymbol{x}_n$ 达到极小,其中 $\boldsymbol{x}_j$ 是 n_j 维列向量,$\boldsymbol{b}_j$ 是 m_j 维列向量,$\boldsymbol{b}$ 是 m 维列向

量，B_j 是 $m_j \times n_j$ 阶矩阵，A_j 是 $m \times n_j$ 阶矩阵，$\boldsymbol{c}_j$ 是 n_j 维行向量．也可以把约束条件分开写，即求 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，满足

$$\sum_{j=1}^{n} A_j \boldsymbol{x}_j = \boldsymbol{b}$$

和 $B_j \boldsymbol{x}_j = \boldsymbol{b}_j, \boldsymbol{x}_j \geqslant 0 \ (j=1,2,\cdots,n)$，使

$$\min f = \sum_{j=1}^{n} \boldsymbol{c}_j \boldsymbol{x}_j.$$

丹齐克-沃尔夫分解算法(Dantzig-Wolfe decomposition algorithm)　求解可分解的大规模线性规划问题的一种算法．对于可分解的线性规划问题(称为母规划)，可以分解成几个规模较小的子规划．分解算法的过程是从母规划的一个基可行解开始，作对应的乘数，并将母规划分解成几个子规划．通过解几个子规划来判断这一基可行解是否是最优的．若不是最优的，就利用单纯形法对母规划进行换基迭代，得到一个新的基可行解，再作相应的乘数……，经有限次计算就可以得到母规划的最优解．这种把一个大规模的线性规划问题分解成几个有关系的规模较小的规划问题来计算的方法，最早是福特(Ford，L. R.)和富尔克森(Fulkerson，D. R.)在解多种商品网络流时提出来的．丹齐克(Dantzig，G. B.)和沃尔夫(Wolfe，P.)在福特和富尔克森等人的工作基础上提出求线性规划问题的分解算法．现在，分解原理已经在最优化的很多领域中得到推广，并且成为解决大系统最优化问题的有力工具．

整 数 规 划

整数规划(integer programming)　数学规划的一个分支．是研究决策变量全部或部分取整数值的数学规划问题．整数规划与组合最优化密切相关(参见"组合最优化")．它研究的特殊问题有：背包问题、固定费用问题、送货问题等．

整数线性规划问题(integer linear programming problem)　一类线性规划问题．指决策变量取整数的线性规划．整数规划主要是研究整数线性规划问题．

纯整数规划(pure integer programming)　亦称全整数规划．一类整数规划．指所有决策变量都限制为整数的整数规划问题．

全整数规划(all integer programming)　即"纯整数规划"．

混合整数规划(mixed integer programming)　一类整数规划．指部分决策变量限制为整数的整数规划问题．

0-1 **整数规划**(zero-one integer programming)　一类整数规划．指全部决策变量只取 0 或 1 的整数规划问题．

整数规划的数学模型(mathematical model of integer programming)　整数线性规划问题的数学表达形式．即

$$\begin{cases} \max \ \sum_{j=1}^{n} c_j x_j, \\ \text{s.t.} \ \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i=1,2,\cdots,m), \\ \quad x_j \geqslant 0 \qquad (j=1,2,\cdots,n), \\ \quad x_j \text{ 全为整数或部分为整数}, \end{cases}$$

其中 $x_1, x_2, \cdots, x_n$ 是一组决策变量．

整数规划的解法(method for solving integer programming)　求解整数规划问题的方法．主要有：割平面法(参见"割平面法")、分支定界法(参见"分支定界法")和求解 0-1 整数规划的隐枚举法(参见"隐枚举法")．

割平面法(cutting plane method)　主要用于求解整数线性规划问题的一种方法．1958 年，首先由格莫理(Gomory，R. E.)提出．基本思路是：先不考虑整数性约束，求解相应的线性规划问题．若线性规划问题的最优解恰好是整数解，则此解即为整数规划问题的最优解．否则，就增加一个新的约束条件，称为割平面．割平面必须具有两条性质：

1. 从线性规划问题的可行域中至少割掉目前的非整数最优解．

2. 不割掉任何整数可行域，然后在缩小的可行域上继续解线性规划问题．

重复以上作法，经有限次切割后，必可在缩小的可行域的一个整数极点上达到整数规划问题的最优解．

具体用于求解的割平面有整数型割平面、对偶-全整数割平面、原-全整数割平面等．

分支定界法(branch and bound method)　求解整数规划问题的一个重要而有效的方法．指经戴肯(Daikin，R. J.)改进的方法．基本思路是：先不考虑整数性约束，解相应的线性规划问题．若线性规划问题的最优解满足整数性约束，则它就是整数规划问题的最优解．否则，就在不满足整数性约束的变量中，任选一个 x_i，设它的值为 b_{i0}，分别增加约束条件

$$x_i \leqslant [b_{i0}] \text{ 和 } x_i \geqslant [b_{i0}] + 1,$$

于是线性规划问题的可行域被分成两个子域，中间割去了

$$[b_{i0}] < x_i < [b_{i0}] + 1$$

部分，在这部分内不含有原整数规划问题的解．此时，原问题分解为两个子问题，称其为分支，它们相应的线性规划问题记为(1)和(2)．分别求解(1)和(2)，对于不满足原整数约束的问题，若需要，则按上

述方法继续寻找新的分支，再求解它们相应的线性规划问题，直到不需要再分支. 于是，在求得的整数解中找到原整数规划问题的最优解. 在分支的同时，相应地确定了原问题的上界与下界，分支定界法因此而得名. 在求解过程中，有些可行解已被全部隐含枚举了，因此分支定界法是一种隐枚举法(参见“隐枚举法”). 分支定界法不仅适用于解纯整数规划问题，对于混合整数规划问题也是有效的，而且对于约束条件多的大型整数线性规划问题更有其优越性.

枚举法(enumeration method)　亦称穷举法. 求解整数规划问题的一种方法. 是把问题的所有解(组合)列出，从中定出最优解的方法.

隐枚举法(implicit enumeration method)　一种特殊的分支定界法. 对 0-1 规划问题. 利用变量只能取 0 或 1 的两个值的特性，进行分支定界，以达到隐枚举的目的. 基本思路是：通过变量的变换，使目标函数中的系数全为非正. 首先令全部变量取 0，因为目标函数的系数全非正，所以此解相应的目标函数值 $s=0$ 就是上界. 若可行，则此解为最优解，计算终止. 否则，有选择地指定其中某个变量为 0 或 1，并把它们固定下来(称为固定变量)，将问题分解成两个子问题. 然后，分别对它们进行检验，即对未被固定取值的变量(称为自由变量)，令其全部为 0，检查它们与固定变量所组成的解是否可行. 若可行，则此解就是目前最好的可行解(不一定是最优解)，不再分支，其相应的目标函数值就是原问题的一个下界；否则，在余下的自由变量中，继续上面的过程. 经过检验，或者停止分支，修改下界，或者有选择地将某个自由变量转为固定变量，指定其为 0 或 1，把子问题再分支. 如此进行下去，直到全部子问题停止分支，或没有自由变量为止，而以其中最大的下界值所对应的可行解为最优解.

过滤条件(filtering condition)　一种约束条件. 求解 0-1 整数规划时，作为新的约束条件增加到原问题中的约束条件.

0-1 **整数规划的解法**(solving method for zero-one integer programming)　求解 0-1 整数规划的方法. 主要有两种：

1. 隐枚举法(参见“隐枚举法”)，在决策变量较多的情况下采用.

2. 枚举法的改进(参见“枚举法”)，在决策变量较少的情况下采用. 这种方法的基本思路是：首先将全部决策变量取 0 或 1 的所有解(组合)列出，然后在逐个检查这些解是否可行的过程中，利用增加并不断修改过滤条件(参见“过滤条件”)的办法，以达到减少计算量、求得最优解的目的.

背包问题(knapsack problem)　一种特殊的 0-1 整数规划问题. 数学模型为

$$\begin{cases} \max\ \sum_{j=1}^{n} c_j x_j, \\ \text{s.t.}\ \sum_{j=1}^{n} a_j x_j \leqslant b, \\ \qquad x_j \text{ 为 } 0 \text{ 或 } 1 \quad (j=1,2,\cdots,n). \end{cases}$$

它可以解释为一位旅行者在出发前必须决定携带哪些物品，这里的 c_j 表示携带第 j 种物品的“价值”或效益，其重 $a_j>0$(千克)；携带各种物品的总重量不应超过 b 千克的限制；目标是使旅途中所携带物品的总“价值”或总效益最大. 经济活动中有许多整数规划问题与它等价，故把这类问题通称为背包问题，亦称投资问题(参见“投资问题”).

投资问题(investment problem)　一种特殊的 0-1 整数规划问题. 数学模型为

$$\begin{cases} \max\ \sum_{j=1}^{n} c_j x_j, \\ \text{s.t.}\ \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i=1,2,\cdots,m), \\ \qquad x_j = 0 \text{ 或 } 1 \quad (j=1,2,\cdots,n). \end{cases}$$

它可以解释为有 n 个投资项目可供选择，这里的 c_j 表示从第 j 项投资可得到的收益，a_{ij} 表示第 i 种资源用于第 j 项投资的数量，b_i 表示第 i 种资源的限量，目标是在各种资源不超过限量 b_i 的条件下，对哪些项目投资，才使整个投资的总效益为最大？当问题中仅有一个不等式约束条件时，它就是背包问题(参见“背包问题”).

投资选择问题(selection problem of investment item)　一类特殊的 0-1 整数规划问题. 指将一定数量的资源分配到不同的投资项目中去，在若干种可行的投资方案中做出选择，以使投资的效益最大的问题. 用 $x_j=0$ 表示不对第 j 个项目进行投资，$x_j=1$ 表示对第 j 个项目进行投资，其数学模型参见“投资问题”. 若在投资的 n 个项目中有 k 个是相互排斥的，则在“投资问题”的数学模型中添加新的约束条件

$$\sum_{j=1}^{k} x_j \leqslant 1,$$

它表示在头 k 个项目投资中至多只能选择一项. 若某项新投资是依赖于先前的投资而定的，则在“投资问题”的数学模型中加入新的约束条件：$x_j \geqslant x_i$，它表示若 $x_i=1$，即选择第 i 个项目，则必定有 $x_j=1$，即一定要选择第 j 个项目. 若问题是要求在前面的两个约束条件中满足其中的一个，而不是两个约束条件都要满足，则可将原约束的前两个修改为

$$\sum_{j=1}^{n} a_{1j} x_j \leqslant b_1 + yM,$$

$$\sum_{j=1}^{n} a_{2j}x_j \leqslant b_2 + (1-y)M,$$

其中 y 为新引进的 0-1 变量，M 为充分大的正数. 若问题是要求在 m 个约束条件中至少满足 k 个($1\leqslant k\leqslant m$)，则可将原约束条件修改为

$$\begin{cases} \sum_{j=1}^{n} a_{ij}x_j \leqslant b_i + (1-y_i)M \quad (i=1,2,\cdots,m), \\ \sum_{i=1}^{m} y_i \geqslant k, \\ \text{所有的 } x_j, y_i \text{ 为 0 或 1}, \end{cases}$$

其中 M 为充分大的正数，k 为整数.

仓库选用问题(selection problem of store)　一类混合整数规划问题. 决策者要在 n 个仓库中决定选用哪几个，以满足 m 个顾客对货物的需要，即每个顾客的需要量 $b_j(j=1,2,\cdots,m)$ 必须从各仓库中得到满足，且只能从动用的仓库中得到满足，而对动用的仓库必须支付固定的费用(如租金、管理费等)；同时，还应决定从动用的哪些仓库中运多少货物到顾客处，使总的耗费为最小. 数学模型为：求一组变量 x_{ij} 和 y_i，使

$$\begin{cases} \min \sum_{i=1}^{n}\left[g_i y_i + \sum_{j=1}^{m} c_{ij}x_{ij}\right], \\ \text{s.t.} \ \sum_{i=1}^{n} x_{ij} = b_j \quad (j=1,2,\cdots,m), \\ \quad \sum_{j=1}^{n} x_{ij} \leqslant y_i d_i \quad (i=1,2,\cdots,n), \\ \quad \text{所有的 } x_{ij} \geqslant 0, \text{所有的 } y_i = 0 \text{ 或 } 1, \end{cases}$$

其中 x_{ij} 表示从仓库 i 运送给顾客 j 的货物量. $y_i=1$ 表示仓库 i 被动用，$y_i=0$ 表示仓库 i 不被动用，g_i 表示当使用仓库 i 时的固定费用，d_i 表示仓库 i 的允许容量，c_{ij} 表示从仓库 i 运送货物到顾客 j 处的单位运费. 由问题中的不等式约束条件，对于某个具体的 i 而言，当 $x_{ij}>0$ 时，y_i 只能为 1，即第 i 个仓库被动用，其费用为

$$g_i + \sum_{j=1}^{m} c_{ij}x_{ij};$$

当 $x_{ij}=0$ 时，y_i 可能为 0 或 1，但结合目标函数是求最小值，因而 $y_i=0$.

非线性规划

非线性规划(nonlinear programming)　数学规划的一个重要分支. 它研究目标函数或约束条件中的函数有一个或多个是变量的非线性函数的数学规划问题. 其研究的问题，称为非线性规划问题，简称非线性规划，记为(NP). 极小化形式的非线性规划问题的数学模型为

$$\begin{cases} \min f(\boldsymbol{x}), \\ \text{s.t.} \ g_i(\boldsymbol{x}) \leqslant 0 \quad (i=1,2,\cdots,p), \\ \quad h_j(\boldsymbol{x}) = 0 \quad (j=1,2,\cdots,q), \end{cases}$$

其中 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 是 n 维欧氏空间 R^n 中的向量，$f(\boldsymbol{x})$ 是目标函数，$g_i(\boldsymbol{x})\leqslant 0(i=1,2,\cdots,p)$ 和 $h_j(\boldsymbol{x})=0(j=1,2,\cdots,q)$ 是约束条件，并且在 $f(\boldsymbol{x})$，$g_i(\boldsymbol{x})(i=1,2,\cdots,p)$ 和 $h_j(\boldsymbol{x})(j=1,2,\cdots,q)$ 中至少有一个是 $\boldsymbol{x}$ 的非线性函数. 因 q 个等式约束 $h_j(\boldsymbol{x})=0(j=1,2,\cdots,q)$ 可以化成与它等价的 $2q$ 个不等式约束

$$\begin{cases} h_j(\boldsymbol{x}) \leqslant 0, \\ -h_j(\boldsymbol{x}) \leqslant 0 \quad (j=1,2,\cdots,q), \end{cases}$$

故非线性规划的数学模型可写成另一形式

$$\min_{\boldsymbol{x}\in X} f(\boldsymbol{x}),$$

其中 $X=\{\boldsymbol{x} \mid g_i(\boldsymbol{x})\leqslant 0, i=1,2,\cdots,p\}$.

等值线(contour line)　数学规划中的概念. 指数学规划问题中，目标函数在其上取值相等的一条曲线. 因为令 $f(\boldsymbol{x})=c$(c 为常数)，则它一般地表示为一条曲线(或一张曲面)，故称为等值线(或等值面).

等值面(contour surface)　见“等值线”.

梯度向量(gradient vector)　一种特殊向量. 可微多元函数的各偏导数组成的向量. 设 f 为 R^n 上的一阶可微函数，则它在点 $\bar{\boldsymbol{x}}\in\mathrm{R}^n$ 处的梯度向量是

$$f_x(\bar{\boldsymbol{x}}) = \nabla f(\bar{\boldsymbol{x}}) = \left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_1}, \frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_2}, \cdots, \frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_n}\right)^{\mathrm{T}},$$

其中 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$.

黑塞矩阵(Hessian matrix)　一种特殊矩阵. 二阶可微多元函数的各二阶偏导数组成的矩阵. 设 f 为 R^n 上的二阶连续可微函数，则它在点 $\bar{\boldsymbol{x}}\in\mathrm{R}^n$ 处的黑塞矩阵是

$$H(\bar{\boldsymbol{x}}) = f_{xx}(\bar{\boldsymbol{x}}) = \nabla^2 f(\bar{\boldsymbol{x}}) = \begin{pmatrix} \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_1 \partial x_1} & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_n \partial x_n} \end{pmatrix},$$

其中 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$.

正定二次函数(positive definite quadratic function)　系数矩阵为对称正定矩阵的二次函数. 设 $\boldsymbol{x}\in\mathrm{R}^n$，$A$ 为 $n\times n$ 对称正定矩阵，$\boldsymbol{b}\in\mathrm{R}^n$ 为常向量，c 为常数，则二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c$$

称为正定二次函数.

局部极值(local extreme value)　实值函数在

某点邻域内取得极小或极大的值. 设 X 为 $\mathbb{R}^n$ 中的集合, f 为 X 上的实值函数, 对于 $\boldsymbol{x}^* \in X$, 若存在某个 $\varepsilon>0$, 使所有与 $\boldsymbol{x}^*$ 的距离小于 ε 的 $\boldsymbol{x}\in X$ 均满足不等式 $f(\boldsymbol{x}^*)\leqslant f(\boldsymbol{x})$ ($f(\boldsymbol{x}^*)\geqslant f(\boldsymbol{x})$), 则称 $\boldsymbol{x}^*$ 为 $f(\boldsymbol{x})$在 X 上的一局部极小(大)点, $f(\boldsymbol{x}^*)$为局部极小(大)值. 若对于所有的 $\boldsymbol{x}\neq\boldsymbol{x}^*$, 且与 $\boldsymbol{x}^*$ 的距离小于 ε 的 $\boldsymbol{x}\in X$ 有 $f(\boldsymbol{x}^*)<f(\boldsymbol{x})$ ($f(\boldsymbol{x}^*)>f(\boldsymbol{x})$), 则称 $\boldsymbol{x}^*$ 为 $f(\boldsymbol{x})$在 X 上的一严格局部极小(大)点, $f(\boldsymbol{x}^*)$为严格局部极小(大)值.

全局极值(global extreme value) 实值函数在某区域取得极小或极大的值. 设 X 为 $\mathbb{R}^n$ 中的集合, f 为 X 上的实值函数, 对于 $\boldsymbol{x}^*\in X$, 如果对所有的 $\boldsymbol{x}\in X$ 均满足 $f(\boldsymbol{x}^*)\leqslant f(\boldsymbol{x})$ ($f(\boldsymbol{x}^*)\geqslant f(\boldsymbol{x})$), 则称 $\boldsymbol{x}^*$ 为 $f(\boldsymbol{x})$在 X 上的全局极小(大)点, $f(\boldsymbol{x}^*)$为全局极小(大)值. 若对所有的 $\boldsymbol{x}\in X$, 且 $\boldsymbol{x}\neq\boldsymbol{x}^*$, 均有$f(\boldsymbol{x}^*)<f(\boldsymbol{x})$ ($f(\boldsymbol{x}^*)>f(\boldsymbol{x})$), 则称 $\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$在 X 上的严格全局极小(大)点, $f(\boldsymbol{x}^*)$为严格全局极小(大)值.

总体极值(global extreme value) 即"全局极值". 总体极小(大)点即全局极小(大)点, 总体极小(大)值即全局极小(大)值.

极值的必要条件(necessary condition of extreme value) 函数在局部极小点或局部极大点处必须满足的条件. 设 $\boldsymbol{x}^*$ 是集合 $X\subset\mathbb{R}^n$ 的内点, f 是定义在 X 上的实值函数. 若 f 在 $\boldsymbol{x}^*$ 处可微, 且在该点处有局部极小值或局部极大值, 此时 $\nabla f(\boldsymbol{x}^*)=\boldsymbol{0}$ 即为一必要条件.

极值的充分条件(sufficient condition of extreme value) 使函数在一点处达到局部极小值或局部极大值的条件. 设 $\boldsymbol{x}^*$ 是集合 $X\subset\mathbb{R}^n$ 的内点, f 是定义在 X 上的实值函数, 且在该点处二阶连续可微. 若 $\nabla f(\boldsymbol{x}^*)=\boldsymbol{0}$, 且对任何非零向量 $\boldsymbol{z}$, 有 $\boldsymbol{z}^{\mathrm{T}}H(\boldsymbol{x}^*)\boldsymbol{z}>0$, 则 $f(\boldsymbol{x})$在 $\boldsymbol{x}^*$ 处有严格局部极小值. 若$\boldsymbol{z}^{\mathrm{T}}H(\boldsymbol{x}^*)\boldsymbol{z}<0$, 则 $f(\boldsymbol{x})$在 $\boldsymbol{x}^*$ 处有严格局部极大值. 这里 $H(\boldsymbol{x}^*)$为 f 在 $\boldsymbol{x}^*$ 处的黑塞矩阵.

拉格朗日鞍点(Lagrange saddle point) 非线性规划问题中满足特定条件的点. 对于非线性规划问题(NP)(参见"非线性规划"), 它的拉格朗日函数是指目标函数和约束条件中函数的如下线性组合:

$$L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{p}\lambda_i g_i(\boldsymbol{x}) + \sum_{j=1}^{q}\mu_j h_j(\boldsymbol{x}),$$

其中 $\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_p)^{\mathrm{T}}\geqslant\boldsymbol{0}$, $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_q)^{\mathrm{T}}$. 满足条件

$$L(\boldsymbol{x}^*,\boldsymbol{\lambda},\boldsymbol{\mu}) \leqslant L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*) \leqslant L(\boldsymbol{x},\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$$
$$(\boldsymbol{x},\boldsymbol{x}^* \in \mathbb{R}^n;\boldsymbol{\lambda},\boldsymbol{\lambda}^* \in \mathbb{R}^p,\boldsymbol{\lambda}\geqslant 0,\boldsymbol{\lambda}^*\geqslant\boldsymbol{0};$$
$$\boldsymbol{\mu},\boldsymbol{\mu}^* \in \mathbb{R}^q)$$

的点$(\boldsymbol{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$称为(NP)的拉格朗日鞍点.

凸函数(convex function) 凸集上满足特定条件的一类函数. 设 S 是线性空间中的凸集, f 是 S 上的实值函数. 若对任何实数 $\alpha\in(0,1)$和 S 中的任意两点 $\boldsymbol{x}^1$ 和 $\boldsymbol{x}^2$, 恒有 $f(\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2)\leqslant\alpha f(\boldsymbol{x}^1)+(1-\alpha)f(\boldsymbol{x}^2)$, 则称 f 是 S 上的凸函数. 若对 $\alpha\in(0,1)$和任意两个不同点 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$, 恒有 $f(\alpha\boldsymbol{x}^1+(1-\alpha)\cdot\boldsymbol{x}^2)<\alpha f(\boldsymbol{x}^1)+(1-\alpha)f(\boldsymbol{x}^2)$, 则称 f 是 S 上的严格凸函数. 在 20 世纪初, 赫尔德(Hölder, O. L.)和延森(Jensen, J. L. W. V.)奠定了凸函数的理论基础.

凸函数的主要性质有:

1. 若 f 为定义在凸集 S 上的凸函数, 则对任意实数 $\beta\geqslant0$, 函数 βf 也是定义在 S 上的凸函数.

2. 若 f_1 和 f_2 为定义在凸集 S 上的两个凸函数, 则其和 $f=f_1+f_2$ 仍为定义在 S 上的凸函数.

3. 若 $f_i(i=1,2,\cdots,m)$为定义在凸集 S 上的凸函数, 则对任意实数 $\beta_i\geqslant0$, 函数

$$\sum_{i=1}^{m}\beta_i f_i$$

也是定义在 S 上的凸函数.

4. 若 f 为定义在凸集 S 上的凸函数, 则对每一实数 c, 水平集 $S_c=\{\boldsymbol{x}\mid\boldsymbol{x}\in S,f(\boldsymbol{x})\leqslant c\}$是凸集.

凹函数(concave function) 与凸函数相对应的一类函数. 设 S 是线性空间中的凸集. 若 $-f$ 是 S 上的凸函数, 则称 f 是 S 上的凹函数. 若 $-f$ 是 S 上的严格凸函数, 则称 f 是 S 上的严格凹函数.

支撑集(support set) 一类点集. $\mathbb{R}^n$ 上凸函数 f 的支撑集为 $\mathbb{R}^{n+1}$中的如下点集: $L(f)=\{(\boldsymbol{a},b)\mid\boldsymbol{a}\in\mathbb{R}^n,b\in\mathbb{R},\forall\boldsymbol{x}\in\mathbb{R}^n$ 有 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}-b\leqslant f(\boldsymbol{x})\}$. 对于 $\mathbb{R}^n$ 中的线性函数 $h(\boldsymbol{x})=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}-b$, 若对任意的 $\boldsymbol{x}$ 有 $h(\boldsymbol{x})\leqslant f(\boldsymbol{x})$, 则称 h 为 f 的线性支撑函数.

凸包(convex hull) 由凸组合构成的集合. 设 S 是线性空间中的集合, 称 S 的任意个点的凸组合所构成的集合为 S 的凸包(参见"凸组合").

函数的闭包(closure of function) 函数的线性支撑函数的上确界. 函数 $f(\boldsymbol{x})$的闭包为

$$\mathrm{cl}\,f(\boldsymbol{x}) = \sup_{(\boldsymbol{a},b)\in L(f)}\{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}-b\}.$$

对一切 $\boldsymbol{x}\in\mathbb{R}^n$, 有 $\mathrm{cl}\,f(\boldsymbol{x})\leqslant f(\boldsymbol{x})$. 若 $f=\mathrm{cl}\,f$, 则称凸函数 f 为闭的. 其充分必要条件是: 对每个实数 c, 水平集$\{\boldsymbol{x}\mid\boldsymbol{x}\in\mathbb{R}^n,f(\boldsymbol{x})\leqslant c\}$是闭的.

凸函数的连续性(continuity of convex function) 凸函数的性质. 若 f 是定义在开凸集 S 上的凸函数, 则 f 在 S 上必为连续函数.

判定函数凸性的条件(decision condition of convex function) 凸函数应具备的条件. 有一阶条件和二阶条件:

1. 一阶条件. 设定义在 n 维欧氏空间中某开凸

集 S 上的函数 f 具有连续一阶导数，则 f 为 S 上的凸函数的充分必要条件是：对任意两点 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S,\boldsymbol{x}^1\neq\boldsymbol{x}^2$，恒有 $f(\boldsymbol{x}^2)\geqslant f(\boldsymbol{x}^1)+\nabla f(\boldsymbol{x}^1)^{\mathrm{T}}(\boldsymbol{x}^2-\boldsymbol{x}^1)$（如果此式为严格不等式，它就是严格凸函数的充分必要条件）.

2. 二阶条件. 设定义在 n 维欧氏空间中某一开凸集 S 上的函数 f 具有连续二阶导数，则 f 为 S 上的凸函数的充分必要条件是：f 的黑塞矩阵 $H(\boldsymbol{x})$ 在 S 上半正定. 若 f 的黑塞矩阵对一切 $\boldsymbol{x}\in S$ 都是正定的，则 f 是 S 上的严格凸函数.

凸函数的极值（extreme value of convex function） 凸函数在某点邻域（或某区域）内取得的极小值或极大值. 凸函数的极值有以下性质：

1. 若 f 为定义在凸集 S 上的凸函数，则它的任一局部极小点就是它在 S 上的全局极小点，而且它的极小点形成一个凸集.

2. 设 f 为定义在凸集 S 上的可微凸函数，若存在点 $\boldsymbol{x}^*\in S$，使得对于所有的 $\boldsymbol{x}\in S$，有 $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\cdot(\boldsymbol{x}-\boldsymbol{x}^*)\geqslant 0$，则 $\boldsymbol{x}^*$ 是 f 在 S 上的全局极小点.

3. 定义在凸集上的凸函数的驻点（梯度为 $\mathbf{0}$ 的点），就是其全局极小点. 全局极小点并不是惟一的，但若为严格凸函数，则其全局极小点是惟一的.

凸规划（convex programming） 约束集为凸集，目标函数为（凸约束集上的）凸函数的非线性规划（问题）. 设 f 和 $g_i(i=1,2,\cdots,p)$ 为凸函数，$h_j(j=1,2,\cdots,q)$ 为线性函数，则称非线性规划（问题）

$$\begin{cases}\min\limits_{\boldsymbol{x}\in X} f(\boldsymbol{x}),\\ X=\{\boldsymbol{x}\mid g_i(\boldsymbol{x})\leqslant 0,i=1,2,\cdots,p;\\ \qquad h_j(\boldsymbol{x})=0,j=1,2,\cdots,q\}\end{cases}$$

为凸规划（问题）. 这时 X 为凸集.

凸规划的可行域（feasible region of convex programming） 一类凸集. 指满足凸规划问题中约束条件的所有点组成的集合.

凸规划的最优解（optimal solution of convex programming） 使目标函数达到最优值的凸规划问题的解. 凸规划的局部最优解即为全局最优解. 当凸规划的目标函数 $f(\boldsymbol{x})$ 为严格凸函数时，若最优解存在，则最优解必定惟一. 最优解的集合形成一个凸集.

拟凸函数（quasi-convex function） 凸集上的一类函数. 设 S 是线性空间中的非空凸集，f 是 S 上的实值函数. 若对任何实数 $\alpha\in(0,1)$ 和 S 中的任意两点 $\boldsymbol{x}^1$ 和 $\boldsymbol{x}^2$，恒有

$$f(\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2)\leqslant\max\{f(\boldsymbol{x}^1),f(\boldsymbol{x}^2)\},$$

则称 f 是 S 上的拟凸函数，或 f 在 S 上是拟凸的. S 上的凸函数也是 S 上的拟凸函数. 对于任何 $\alpha\in(0,1)$ 和任意的 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$，若 $f(\boldsymbol{x}^2)\geqslant f(\boldsymbol{x}^1)$，而 $f(\boldsymbol{x}^2)$ 不小于在 $\boldsymbol{x}^1$ 与 $\boldsymbol{x}^2$ 连线上一切点的函数值，则 f 是拟凸的.

严格拟凸函数（strictly quasi-convex function） 凸集上的一类函数. 设 S 是线性空间中的非空凸集，f 是 S 上的实值函数. 若对任何实数 $\alpha\in(0,1)$ 和任意的 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$，且 $f(\boldsymbol{x}^1)\neq f(\boldsymbol{x}^2)$，恒有

$$f(\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2)<\max\{f(\boldsymbol{x}^1),f(\boldsymbol{x}^2)\},$$

则称 f 是 S 上的严格拟凸函数，或 f 在 S 上是严格拟凸的.

强拟凸函数（strongly quasi-convex function） 凸集上的一类函数. 设 S 是线性空间中的非空凸集，f 是 S 上的实值函数. 若对任何实数 $\alpha\in(0,1)$ 和任意的 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$，且 $\boldsymbol{x}^1\neq\boldsymbol{x}^2$，恒有

$$f(\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2)<\max\{f(\boldsymbol{x}^1),f(\boldsymbol{x}^2)\},$$

则称 f 为 S 上的强拟凸函数，或 f 在 S 上是强拟凸的. 每个严格凸函数都是强拟凸函数. 每个强拟凸函数都是严格拟凸函数.

拟凸函数的充分必要条件（necessary and sufficient condition of quasi-convex function） 拟凸函数应满足的条件. f 是 S 上的拟凸函数的充分必要条件是：对每一实数 c，水平集 $S_c=\{\boldsymbol{x}\mid\boldsymbol{x}\in S,f(\boldsymbol{x})\leqslant c\}$ 是凸集.

拟凹函数（quasi-concave function） 与拟凸函数相对应的一类函数. 设 $-f$ 是凸集 S 上的拟凸函数，称 f 是 S 上的拟凹函数. 若 $f(\boldsymbol{x}^2)\geqslant f(\boldsymbol{x}^1)$，且 $\boldsymbol{x}^1$ 与 $\boldsymbol{x}^2$ 连线上一切点的函数值都大于或等于 $f(\boldsymbol{x}^1)$，则 f 是拟凹的. 若 $-f$ 是严格拟凸函数，则称 f 是严格拟凹函数. 若 $-f$ 是强拟凸函数，则称 f 是强拟凹函数.

伪凸函数（pseudo-convex function） 凸集上的一类函数. 设 $S\subset\mathrm{R}^n$ 是凸集，f 是 S 上的可微函数. 若对任意的 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$，以及 $\nabla f(\boldsymbol{x}^1)^{\mathrm{T}}(\boldsymbol{x}^2-\boldsymbol{x}^1)\geqslant 0$，有 $f(\boldsymbol{x}^2)\geqslant f(\boldsymbol{x}^1)$，则称 f 是 S 上的伪凸函数. 或者等价地，若 $f(\boldsymbol{x}^2)<f(\boldsymbol{x}^1)$，有 $\nabla f(\boldsymbol{x}^1)^{\mathrm{T}}(\boldsymbol{x}^2-\boldsymbol{x}^1)<0$，则 f 是 S 上的伪凸函数. 若对任意的 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S,\boldsymbol{x}^1\neq\boldsymbol{x}^2$，以及 $\nabla f(\boldsymbol{x}^1)^{\mathrm{T}}(\boldsymbol{x}^2-\boldsymbol{x}^1)\geqslant 0$，有 $f(\boldsymbol{x}^2)>f(\boldsymbol{x}^1)$，或者等价地，若 $f(\boldsymbol{x}^2)\leqslant f(\boldsymbol{x}^1)$，则有 $\nabla f(\boldsymbol{x}^1)^{\mathrm{T}}(\boldsymbol{x}^2-\boldsymbol{x}^1)<0$，则称 f 是 S 上的严格伪凸函数.

伪凸函数与拟凸函数之间有下述关系：若 f 是 $S\subset\mathrm{R}^n$（非空开凸集）上的伪凸函数，这时 f 既是严格拟凸函数也是拟凸函数. 若 f 是 S 上的严格伪凸函数，则 f 是 S 上的强拟凸函数.

伪凹函数（pseudo-concave function） 与伪凸函数相对应的一类函数. 设 $-f$ 是凸集 $S\subset\mathrm{R}^n$ 上的伪凸函数，则称 f 是 S 上的伪凹函数. 若 $-f$ 是严格伪凸函数，则称 f 是严格伪凹函数.

全局极小点(global minimum point)　在可行域 $X \subset R^n$ 上使目标函数 $f(\boldsymbol{x})$ 达到极小值的点. 即:

1. 设 f 是定义在开凸集 $X \subset R^n$ 上的连续拟凸函数, 且在 $\boldsymbol{x}^* \in X$ 处可微, $\nabla f(\boldsymbol{x}^*)^T = \boldsymbol{0}$, 则 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 在 X 上的全局极小点的充分必要条件是: 对任意的 $\boldsymbol{x} \in X$, 恒有 $\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x}-\boldsymbol{x}^*) \geqslant 0$.

2. 设 f 是开凸集 $X \subset R^n$ 上的伪凸函数, 又设对某个点 $\boldsymbol{x}^* \in X$ 有 $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$, 则 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 在 X 上的全局极小点.

3. 设 f 是凸集 $X \subset R^n$ 上的严格拟凸函数, $\boldsymbol{x}^* \in X$ 是 $f(\boldsymbol{x})$ 在 X 上的局部极小点, 则 $\boldsymbol{x}^*$ 必为 $f(\boldsymbol{x})$ 在 X 上的全局极小点.

4. 设 f 是凸集 X 上的拟凸函数, 若 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 在 X 上的一个严格局部极小点, 则 $\boldsymbol{x}^*$ 也是 $f(\boldsymbol{x})$ 在 X 上的严格全局极小点.

凸锥(convex cone)　一类特殊的凸集. 设 S 是线性空间中的集合. 若对于任意数 $\alpha \geqslant 0, \beta \geqslant 0$ 以及任意的 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$, 都有 $\alpha\boldsymbol{x}^1 + \beta\boldsymbol{x}^2 \in S$, 则称 S 是凸锥. 凸锥一定是凸集.

闭凸锥包(closed convex cones hull)　一种凸锥. 设 S 是线性空间中的集合, 则称包含 S 的最小的闭凸锥为 S 的闭凸锥包, 或称由 S 生成的闭凸锥, 记为 $C(S)$.

上图收敛性(epigraph convergence)　最优化逼近理论中的收敛性质. R^n 上函数 f 的上图是指 R^{n+1} 中的集合: $\text{epi}\, f = \{(\boldsymbol{x}, \alpha) \mid \alpha \geqslant f(\boldsymbol{x})\}$. 若给定 $\text{epi}\, f$, 则可完全确定这个函数, 即 $f(\boldsymbol{x}) = \min\{\alpha \mid (\boldsymbol{x}, \alpha) \in \text{epi}\, f\}$. 因此, 集合 $\text{epi}\, f$ 的性质与函数 f 的性质有着决定性的联系: f 为下半连续函数当且仅当 $\text{epi}\, f$ 为 R^{n+1} 中的闭集; f 为凸函数当且仅当 $\text{epi}\, f$ 为凸集. 根据函数与它的上图之间的一一对应关系, 可用上图序列的收敛性来研究函数序列的收敛性, 而上图序列的收敛性即是集合序列的收敛性.

有如下相关的理论: 如果

$$\liminf S_i \supset S \supset \limsup S_i,$$

称集合序列 $\{S_i\}$ 收敛于 S, 这时记为

$$\lim_{i\to\infty} S_i \stackrel{K}{=} S \text{ 或 } S_i \xrightarrow{K} S,$$

其中 $\liminf S_i = \{\boldsymbol{x} \mid$ 存在 $\{\boldsymbol{x}_i\}$ 使得 $\boldsymbol{x}_i \to \boldsymbol{x}$, 且除了有限多个 i 外都有 $\boldsymbol{x}_i \in S_i\}$, $\limsup S_i = \{\boldsymbol{x} \mid$ 存在子序列 $\{\boldsymbol{x}_{i_k}\}$ 使得 $\boldsymbol{x}_{i_k} \to \boldsymbol{x}$, 且 $\boldsymbol{x}_{i_k} \in S_{i_k}\}$. 对于 $\liminf S_i$ 与 $\limsup S_i$, 下面的关系式成立:

$$\lim\inf S_i = \text{cl}\left(\bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} \text{cl}\, S_i\right),$$

$$\limsup S_i = \bigcap_{k=1}^{\infty} \text{cl}\left(\bigcup_{i=k}^{\infty} S_i\right),$$

式中 $\text{cl}A$ 表示集合 A 的闭包. 对任何集合序列 $\{S_i\}$, 有

$$\text{cl} \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} \text{cl}\, S_i = \bigcap_{H \subset \mathcal{N}} \text{cl}\left(\bigcup_{i\in H} S_i\right),$$

$$\bigcap_{k=1}^{\infty} \text{cl}\left(\bigcup_{i=k}^{\infty} S_i\right) = \bigcap_{H \subset \mathcal{K}} \text{cl}\left(\bigcup_{i\in H} S_i\right),$$

式中 H 为自然数集 $N = \{1, 2, \cdots\}$ 的子集, $\mathcal{K}$ 为 N 的下列形式的子集的全体: $\{k \mid k \in N, k \geqslant n\}$ (n 为某一自然数), $\mathcal{N}$ 为 N 的所有无限子集的全体. 对任一序列 $\{S_i\}$, $\liminf S_i$ 和 $\limsup S_i$ 均为闭集. 对任何集合序列 $\{S_i\}$, 包含关系 $\liminf S_i \subset \limsup S_i$ 成立.

若有 $\limsup(\text{epi}\, f_i) \subset \text{epi}\, f \subset \liminf(\text{epi}\, f_i)$, 则称函数序列 $\{f_i\}$ 上图收敛于 f, 记为 $f_i \xrightarrow{e} f$. 这时,

$$\lim_{i\to\infty}(\text{epi}\, f_i) \stackrel{K}{=} \text{epi}\, f,$$

并称 f 为 $\{f_i\}$ 的 epi 极限. 如果

1. 对任一 $\boldsymbol{x}$ 和任一收敛于 $\boldsymbol{x}$ 的序列 $\{\boldsymbol{x}_i\}$, 有

$$\lim\inf f_i(\boldsymbol{x}_i) \geqslant f(\boldsymbol{x}),$$

2. 对任一 $\boldsymbol{x}$, 总存在收敛于 $\boldsymbol{x}$ 的序列 $\{\boldsymbol{x}_i\}$, 使得

$$\lim\sup f_i(\boldsymbol{x}_i) \leqslant f(\boldsymbol{x}),$$

则函数序列 $\{f_i\}$ 上图收敛于 f.

无约束最优化问题的逼近理论是: 设有一系列极小化问题

$$Z = \min_{\boldsymbol{x}\in R^n} f(\boldsymbol{x}), \quad Z_i = \min_{\boldsymbol{x}\in R^n} f_i(\boldsymbol{x}) \quad (i = 1, 2, \cdots),$$

其中 f_i $(i = 1, 2, \cdots)$ 为一族下半连续函数, 且 $f_i \xrightarrow{e} f$, 则

$$\limsup(\text{Arg}\min f_i) \subset \text{Arg}\min f,$$

其中 $\text{Arg}\min f = \{\boldsymbol{x} \mid f(\boldsymbol{x}) = Z\}$. 若 $\text{Arg}\min f$ 为非空集合, Z 为有限数, 则 $Z \geqslant \limsup Z_i$; 若进一步有收敛序列 $\{\bar{\boldsymbol{x}}_i\}$, $\bar{\boldsymbol{x}}_i \in \text{Arg}\min f_i$, $\bar{\boldsymbol{x}}_i \to \bar{\boldsymbol{x}}$, 则

$$\lim_{i\to\infty} Z_i = Z.$$

设 $f_i \xrightarrow{e} f$, 且存在一个紧致集合 D, 使得对每一 i, $\text{Arg}\min f_i \cap D$ 非空, 则必有 $\lim Z_i = Z$.

约束最优化问题的逼近理论是: 设 $\min f(\boldsymbol{x})$, s.t. $\boldsymbol{x} \in S \subset R^n$, 则它等价于无约束最优化问题

$$\min_{\boldsymbol{x}\in R^n} F(\boldsymbol{x}) = \min_{\boldsymbol{x}\in R^n}[f(\boldsymbol{x}) + \varphi_S(\boldsymbol{x})],$$

式中

$$\varphi_S(\boldsymbol{x}) = \begin{cases} 0 & (\boldsymbol{x} \in S), \\ +\infty & (\boldsymbol{x} \notin S). \end{cases}$$

此问题的近似问题为

$$\begin{cases} \min\ f(\boldsymbol{x}), \\ \text{s.t. } \boldsymbol{x} \in S_i, \end{cases} \tag{1}$$

式中集合序列 $\{S_i\}$ 在某种意义下收敛于 S; 或

$$\begin{cases} \min\ f_i(\boldsymbol{x}), \\ \text{s.t. } \boldsymbol{x} \in S, \end{cases} \tag{2}$$

式中函数序列 $\{f_i\}$ 在某种意义下收敛于 f. 设问题 (1) 与 (2) 中的 f 为连续函数,

$$\lim_{i\to\infty} S_i \stackrel{K}{=} S,$$

则 $F_i(\boldsymbol{x}) \xrightarrow{e} F(\boldsymbol{x})$, 式中,

$$F_i(\boldsymbol{x}) = f(\boldsymbol{x}) + \varphi_{S_i}(\boldsymbol{x}),$$
$$F(\boldsymbol{x}) = f(\boldsymbol{x}) + \varphi_S(\boldsymbol{x}).$$

设 $f_i \to f$，S 为丰满闭集(即 $S=\mathrm{cl}(\mathrm{int}\, S)$)，$f$ 为连续函数，则 $F_i(\boldsymbol{x}) \to F(\boldsymbol{x})$，式中 $F_i(\boldsymbol{x})=f_i(\boldsymbol{x})+\varphi_S(\boldsymbol{x})$，$F(\boldsymbol{x})=f(\boldsymbol{x})+\varphi_S(\boldsymbol{x})$.

上图收敛性是指：若 f 是函数序列 $\{f_i\}$ 的 epi 极限，则 f 一定是下半连续函数. 设 $f_i \xrightarrow{e} f$，且所有 f_i 为凸函数，则 f 亦为凸函数.

凸函数 f 在点 $\bar{x} \in \mathrm{R}^n$ 处的次梯度的集合 $\partial f(\bar{x})$ 称为次微分，定义为 $\partial f(\bar{x})=\{\boldsymbol{v} \in \mathrm{R}^n \mid f(\boldsymbol{x})-f(\bar{\boldsymbol{x}}) \geqslant \boldsymbol{v}(\boldsymbol{x}-\bar{\boldsymbol{x}})\ \forall \boldsymbol{x}\}$，故次微分函数 $\partial f(\bar{x})$ 为集值函数，它的图形 $G(\partial f)$ 定义为 $G(\partial f)=\{(\boldsymbol{x},\boldsymbol{v}) \mid \boldsymbol{v} \in \partial f(\boldsymbol{x})\}$. 有下面的结果：设 $\{f_i\}$ 为凸函数序列，则 $f_i \xrightarrow{e} f$ 当且仅当

$$G(\partial f) = \lim_{i\to\infty} G(\partial f_i).$$

设对任一 $\boldsymbol{x}$，序列 $\{f_i(\boldsymbol{x})\}$ 单调收敛于 $f(\boldsymbol{x})$，式中 f 为下半连续函数，则 $\{f_i\}$ 上图收敛于 f；反之，若 $\{f_i\}$ 为单调序列，$f_i \xrightarrow{e} f$，则 $\{f_i\}$ 亦依点收敛于 f.

无约束最优化问题(unconstrained optimization problem)　不带约束条件的非线性规划问题(参见“非线性规划”).

求无约束最优化问题的迭代过程(iteration method for unconstrained optimization problem)　在无约束最优化问题中，从一个近似解到另一个近似解的迭代过程. 一般迭代过程是：从一近似解(点) $\boldsymbol{x}^k$ 出发，沿着某一选定的方向 $\boldsymbol{p}^k$，选择某一步长 λ_k，得到另一个近似解(点) $\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\lambda_k\boldsymbol{p}^k$，使得 $f(\boldsymbol{x}^{k+1})<f(\boldsymbol{x}^k)$(极小化问题)或 $f(\boldsymbol{x}^{k+1})>f(\boldsymbol{x}^k)$(极大化问题)；当满足某精度要求(如 $\|\nabla f(\boldsymbol{x}^{k+1})\|<\varepsilon$ 或 $\|f(\boldsymbol{x}^{k+1})-f(\boldsymbol{x}^k)\|<\varepsilon$)时，迭代停止. 由于方向 $\boldsymbol{p}^k$ 和步长 λ_k 的不同选择，构成了许多不同的迭代过程.

无约束最优化方法(unconstrained optimization method)　求解无约束最优化问题的方法. 有解析法和直接法两类：

1. 解析法就是利用无约束最优化问题中目标函数 $f(\boldsymbol{x})$ 的解析表达式和它的解析性质(如函数的一阶导数和二阶导数)，给出一种求它的最优解 $\boldsymbol{x}^*$ 的方法，或一种求 $\boldsymbol{x}^*$ 的近似解的迭代方法. 解析法主要有梯度法(或称最速下降法)、共轭方向法、共轭梯度法、非二次函数的共轭梯度法、牛顿法、拟牛顿法、变尺度法等.

2. 直接法就是在求最优解 $\boldsymbol{x}^*$ 的过程中，只用到函数的函数值，而不必利用函数的解析性质. 直接法也是一种迭代法，迭代步骤简单. 当目标函数 $f(\boldsymbol{x})$ 的表达式十分复杂，或写不出具体表达式时，它就成了重要的方法. 一般地，直接法对于目标函数 $f(\boldsymbol{x})$ 只要求能通过逐步试验最后求得近似最优解. 直接法适应面很广，适于计算机运算. 直接法主要有坐标轮换法、爬山法、步长加速法、单纯形调优法、方向加速法等.

搜索方法(search method)　求最优化问题最优解的方法. 对于极小化问题，是指根据目标函数的特征，不断构造搜索方向，逐次使目标函数值下降的迭代方法. 迭代过程是：在极小点的附近选择初始点 $\boldsymbol{x}^0$，按照某种规则确定一搜索方向，在此方向上移动到下一点 $\boldsymbol{x}^k$，并使 $f(\boldsymbol{x}^k)<f(\boldsymbol{x}^0)$；检查所得的新点是否满足所要求的精度，如满足，则停止，$\boldsymbol{x}^k$ 就是所求的近似最优解；若不满足，则在 $\boldsymbol{x}^k$ 处确定新的搜索方向，并在此方向上移动到下一点 $\boldsymbol{x}^{k+1}$，再检查，再移动，直到满足精度要求为止.

一维搜索(one-dimensional search)　亦称线性搜索. 一种搜索方式. 指沿某一已知直线进行寻优的搜索.

线性搜索(linear search)　即“一维搜索”.

序贯试验法(sequential test method)　一维搜索的方法. 为逐步缩短所考察的区间，每一次均根据前面试验的结果安排后面试验的方法.

区间消去法(interval elimination method)　求单变量函数无约束极值的较实用的一类直接搜索方法. 其特点是在搜索过程中，不断缩小最优点所在的区间，即通过搜索区间的逐步缩小来确定最优点.

单峰函数(unimodal function)　在所考虑的区间中只有一个严格局部极大值(峰值)的实值函数(参见“局部极值”). 设 f 是闭区间 $[a,b]$ 上的实值函数. 若存在点 $t^* \in [a,b]$，使得 $f(t)$ 在 $[a,t^*]$ 上严格递增，在 $[t^*,b]$ 上严格递降，则称 f 是 $[a,b]$ 上的单峰函数，$[a,b]$ 是 f 的单峰区间. 若 $-f$ 是 $[a,b]$ 上的单峰函数，则称 f 是 $[a,b]$ 上的单谷函数(下单峰函数)，$[a,b]$ 是 f 的单谷区间.

斐波那契数(Fibonacci numbers)　一种整数. 指数列

$$\{F_n\}=\{F_0,F_1,F_2,F_3,F_4,\cdots\}$$
$$=\{1,1,2,3,5,\cdots\}$$

中的数. 它从第二项开始，每后一个数等于前两个数之和，其递推公式为

$$F_n = F_{n-1} + F_{n-2} \quad (n \geqslant 2).$$

斐波那契(Fibonacci, L.)是 13 世纪的意大利数学家，在他的著作《算盘书》中写了一个有趣的兔子繁殖问题：某人想知道一年内一对兔子可以繁殖成多少对兔子. 他进行了观察，并逐月记录. 设每一对成兔每月生一对幼兔，幼兔经过两个月后成为成兔，开始繁殖. 以 n 表示月份，以 F_n 表示第 n 个月中兔子的对数，他从一对成兔开始，对问题进行了研究. 1753 年，西姆森(Simson, R.)发现 F_{n-1}/F_n 是连分

数

$$\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\cdots}}}$$

的第$(n-1)$个渐近值或渐近分数. 1876 年,吕卡(Lucas,F.-É.-A.)进一步得出斐波那契数的表示式

$$F_n=\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}-\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right]$$
$$(n=0,1,2,\cdots).$$

斐波那契法(Fibonacci method)　亦称斐波那契分数法. 一种一维搜索的区间消去法(参见"区间消去法"). 对闭区间$[a,b]$上的单峰函数$f(t)$,按相邻两斐波那契数之比,使用对称规则进行搜索的方法. 其特点是:逐步缩短所考察的区间,以尽量少的函数求值次数,达到预定的某一缩短率. 设$\{F_n\}$是斐波那契数列,对于区间$[0,1]$,确定搜索次数n,则选择两点:

$$t_1=\frac{F_{n-1}}{F_n},\quad t_2=\frac{F_{n-2}}{F_n},$$

比较$f(t_1)$和$f(t_2)$,消去其中较小值所在的一段区间;在缩短后的搜索区间上,再作第二步迭代;如此继续迭代下去,共做n次迭代,搜索区间长度缩短为$1/F_n$(使小于给定值). 1953 年,美国数学家基弗(Kiefer,J. C.)首先研究用斐波那契数搜索一维函数$f(t)$的局部最大值.

斐波那契分数法(Fibonacci fractional number method)　见"斐波那契法".

黄金分割(golden section)　亦称黄金律,对线段长度的一种分割(方法). 将一线段分割为两部分,使得整个线段之长与分割以后的较长一段的比,等于长的一段与短的一段的比. 设线段长为 1,分割后较长段为ω,短段长为$1-\omega$,则$1:\omega=\omega:(1-\omega)$,由此得到黄金分割数$\omega=0.618033988\cdots$. 圆内接正十边形的一边之长就是圆半径的一个黄金分割.

黄金律(golden rule)　即"黄金分割".

0.618 法(0.618 method)　一种区间消去法(参见"区间消去法"). 是对单峰函数,取搜索区间长度的 0.618(黄金分割数的近似值)倍,按对称规则进行搜索的方法. 每次的试验点均取在区间的0.618(从另一端看是 0.382=1-0.618)倍处. 它以不变的区间缩短率 0.618,代替斐波那契法中每次不同的缩短率. 当$n\to\infty$时,0.618法的缩短率约为斐波那契法的 1.17 倍,故 0.618 法也可以看成是斐波那契法的近似. 0.618法实现起来比较方便,效果也比较好,也是优选法中进行单因素试验常用的方法(参见"优选法").

近似黄金分割法(approximate golden section method)　一种区间消去法(参见"区间消去法"). 利用近似黄金分割数进行一维搜索的方法. 因为黄金分割数ω是无理数,实用中取其某近似值$\tilde{\omega}\approx\omega$作为每次的试验点,按对称规则进行搜索的方法. 0.618 法是近似黄金分割法的重要特例.

成功失败法(success-failure method)　亦称进退法. 一种搜索方法. 为搜索某区间上函数的极小(大)点,每次搜索都要改变搜索步长的一种方法. 如果在第k次迭代沿某方向搜索成功,即函数值下降(上升),下一步仍可沿该方向搜索,而且可以大步向前搜索. 其作法是:从某点t^0出发,步长取为λ. 若$f(t^0+\lambda)<(>)f(t^0)$,则搜索成功,下一步取步长为2λ;如果第n步的步长为$n\lambda$,并搜索成功,下一步取步长为$2n\lambda$. 若在第k次迭代,沿某方向搜索失败,即函数值上升(下降),则应退回原地,下一步沿相反方向,即向后小步搜索. 其作法是:若$f(t^0+\lambda)\geqslant(\leqslant)f(t^0)$,则搜索失败,退回原来点并且再后退$\lambda/4$. 若第$n$步步长为$n\lambda$,搜索失败,则退回到$t^0$后,还要后退$n\lambda/4$. 直到最后搜索步长小于给定的小正数,则停止搜索,得到近似最优点. 这里2λ,$\lambda/4$都是按经验选取的.

进退法(back and forth method)　即"成功失败法".

切线法(tangent line method)　亦称牛顿法. 一种搜索方法. 指在搜索区间上逐次构造新的与所寻求函数相应的二次函数,并且用一系列二次函数的极小点逐步逼近原寻求函数极小点的一种方法. 设$f(t)$是区间$[a,b]$上的单谷函数(参见"单峰函数"),且在所讨论的点$t=t_k$处,能够算出$f(t_k)$,$f'(t_k)$和$f''(t_k)$($f''(t_k)>0$). 构造二次函数

$$\varphi(t)=f(t_k)+f'(t_k)(t-t_k)+\frac{1}{2}f''(t_k)(t-t_k)^2,$$

使它在t_k处的函数值、一阶导数和二阶导数与原寻求函数在t_k处的值相等. 于是,可以用新函数$\varphi(t)$的极小点t_{k+1}作为原寻求函数极小点的近似值,其迭代公式是

$$t_{k+1}=t_k-\frac{f'(t_k)}{f''(t_k)},$$

t_{k+1}与$f(t_k)$的值无关. 若这个近似值不满足预先提出的精度要求,则在t_{k+1}处再构造一个二次函数,并求其极小点,如此逐步逼近原寻求函数$f(t)$的极小点,直至达到某一精度要求为止. 这个方法的收敛速度是相当快的,但要求序列$\{f'(t)\}$收敛于零.

牛顿法(Newton method)　即"切线法".

多项式插值法(polynomial interpolation method)　一种搜索方法. 指用插值多项式$\varphi(t)$的极小

点逼近寻求函数$f(t)$的极小点的方法.具体做法是:求$\varphi'(t)=0$的根,作为$f(t)$的极小点近似,重复应用这一方法进行迭代计算,直到得出满足事先给出的精度要求为止.用二次多项式逼近$f(t)$,称为二次插值法,用三次多项式逼近$f(t)$,称为三次插值法.

抛物线插值法(parabolic interpolation method) 亦称二次插值法.一种多项式插值法(参见"多项式插值法").逐次以拟合的二次曲线的极小点,逼近原寻求函数极小点的一种方法.具体做法是:设$f(t)$在$t_1<t_2<t_3$处的函数值依次为$f(t_1)$,$f(t_2)$和$f(t_3)$,用抛物线$\varphi(t)=a_0+a_1t+a_2t^2$拟合$f(t)$,使满足

$$\begin{cases}\varphi(t_1)=a_0+a_1t_1+a_2t_1^2=f(t_1),\\ \varphi(t_2)=a_0+a_1t_2+a_2t_2^2=f(t_2),\\ \varphi(t_3)=a_0+a_1t_3+a_2t_3^2=f(t_3).\end{cases}$$

对$\varphi(t)$求导并令其等于零,解得$t=-a_1/2a_2$.由上述方程组得到a_1和a_2,将其代入解式便有计算近似极小点的公式:

$$t=\frac{1}{2}\frac{\begin{vmatrix}1 & f(t_1) & t_1^2\\ 1 & f(t_2) & t_2^2\\ 1 & f(t_3) & t_3^2\end{vmatrix}}{\begin{vmatrix}1 & t_1 & f(t_1)\\ 1 & t_2 & f(t_2)\\ 1 & t_3 & f(t_3)\end{vmatrix}}.$$

每次的三点组中,中间点t_2的函数值均不大于搜索区间$[t_1,t_3]$的两端点的函数值.逐次迭代,逐步缩小搜索区间.当相继两次迭代的极小点之间的距离小于某一预先给定的距离时,或者当逼近函数的值和原寻求函数的值之差小于某一允许误差时,即可终止迭代.

三次插值法(cubic interpolation method) 一种多项式插值法(参见"多项式插值法").逐次以三次曲线$\varphi(t)=a_0+a_1t+a_2t^2+a_3t^3$的极小点逼近寻求函数$f(t)$的极小点的一种方法.具体做法是:设$t_1<t_2$,在搜索区间$[t_1,t_2]$中以$\varphi(t)$拟合$f(t)$,使满足

$$\varphi(t_i)=f(t_i),\ \varphi'(t_i)=f'(t_i)\ (i=1,2).\quad(1)$$

对$\varphi(t)$求导并令其等于零,求得$\varphi(t)$在$[t_1,t_2]$中由a_0,a_1,a_2和a_3表示的极小点表达式.再由方程组(1)解得a_0,a_1,a_2和a_3,将其代入此表达式,便得$f(t)$的近似极小点.逐次迭代,当$\varphi(t)$在近似极小点处的导数值的绝对值小于某给定误差时,迭代停止.用三次插值法寻求极小点一般比用二次插值法(参见"抛物线插值法")有更快的收敛速度,但其每一轮迭代的计算量则比二次插值法要大.

负梯度方向(negative gradient direction) 多元可微函数在一点处梯度向量的反方向.设f在R^n中一阶可微,点$\boldsymbol{x}\in\mathrm{R}^n$,则$f$在$\boldsymbol{x}$处的负梯度方向是$-\nabla f(\boldsymbol{x})$.它是使函数$f$在点$\boldsymbol{x}$附近下降最快的方向.

梯度法(gradient method) 亦称最速下降法.求解无约束最优化问题的一种经典而又基本的解析方法.在第$k+1$次迭代点$\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\lambda_k\boldsymbol{p}^k$中,$\boldsymbol{p}^{(k)}=-\nabla f(\boldsymbol{x}^{(k)})$(参见"负梯度方向"),步长$\lambda_k$的选择是多样的.一是$\lambda_k$是否满足不等式$f(\boldsymbol{x}^k-\lambda_k\nabla f(\boldsymbol{x}^k))<f(\boldsymbol{x}^k)$,若满足,继续迭代,否则缩小$\lambda_k$使其满足,这样的$\lambda_k$总是存在的;另一是$\lambda_k$使

$$f(\boldsymbol{x}^k-\lambda_k\nabla f(\boldsymbol{x}^k))=\min_{\lambda\geqslant0}f(\boldsymbol{x}^k-\lambda\nabla f(\boldsymbol{x}^k)),$$

即通过在负梯度方向的一维搜索,来确定使f达到最小的λ_k.确定步长还可以采用其他的一维搜索的方法.

最速下降法(steepest descent method) 即"梯度法".

共轭方向(conjugate direction) 两向量间的一种特殊关系.设A为$n\times n$对称正定矩阵,向量$\boldsymbol{p}^1,\boldsymbol{p}^2\in\mathrm{R}^n$.若满足条件$(\boldsymbol{p}^1)^{\mathrm{T}}A\boldsymbol{p}^2=0$,则称$\boldsymbol{p}^1$和$\boldsymbol{p}^2$关于$A$是共轭方向,或称$\boldsymbol{p}^1$和$\boldsymbol{p}^2$关于$A$共轭.一般地,对于非零向量组$\boldsymbol{p}^1,\boldsymbol{p}^2,\cdots,\boldsymbol{p}^n\in\mathrm{R}^n$,若满足条件:$(\boldsymbol{p}^i)^{\mathrm{T}}A\boldsymbol{p}^j=0(i\neq j,i,j=1,2,\cdots,n)$,则称该向量组关于$A$共轭.

共轭方向法(conjugate direction method) 依次沿共轭方向寻求无约束最优化问题极小点的一类方法.对于n维正定二次函数f,选取关于其系数矩阵是共轭的向量组$\boldsymbol{p}^0,\boldsymbol{p}^1,\cdots,\boldsymbol{p}^{n-1}$,从任一点$\boldsymbol{x}^0\in\mathrm{R}^n$出发,相继以$\boldsymbol{p}^0,\boldsymbol{p}^1,\cdots,\boldsymbol{p}^{n-1}$为搜索方向,迭代公式为:

$$\begin{cases}\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\lambda_k\boldsymbol{p}^k,\\ \lambda_k:\min f(\boldsymbol{x}^k+\lambda\boldsymbol{p}^k).\end{cases}$$

经n次一维搜索,便可找到$\boldsymbol{x}^n$为$f(\boldsymbol{x})$的极小点.共轭方向法是鲍威尔(Powell,M. J. D.)于1964年首先提出的.

鲍威尔法(Powell method) 即"共轭方向法".

共轭梯度法(conjugate gradient method) 共轭方向法中的一种方法.它的搜索方向是利用一维搜索所得极小点处的梯度生成的.对正定二次函数f,共轭梯度法的计算公式为

$$\begin{cases}\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\lambda_k\boldsymbol{p}^k,\\ \lambda_k=-\dfrac{\nabla f(\boldsymbol{x}^k)^{\mathrm{T}}\boldsymbol{p}^k}{(\boldsymbol{p}^k)^{\mathrm{T}}A\boldsymbol{p}^k},\\ \boldsymbol{p}^{k+1}=-\nabla f(\boldsymbol{x}^{k+1})+\beta_k\boldsymbol{p}^k,\\ \beta_k=\dfrac{\|\nabla f(\boldsymbol{x}^{k+1})^{\mathrm{T}}\|^2}{\|\nabla f(\boldsymbol{x}^k)^{\mathrm{T}}\|^2},\\ k=0,1,2,\cdots,n-1,\end{cases}$$

其中$\boldsymbol{x}^0$为初始近似点,$\boldsymbol{p}^0=-\nabla f(\boldsymbol{x}^0)$.此法由弗莱彻(Fletcher,R.)和里夫斯(Reevse,C. M.)提出,亦

称 F-R 共轭梯度法.

非二次函数的共轭梯度法(conjugate gradient method of nonquadratic function)　共轭梯度法的推广. 设 f 是在点 $\boldsymbol{x}^*$ 附近具有二次连续可微的凸函数,$\boldsymbol{x}^0$是初始点,迭代公式为

$$\begin{cases}\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\lambda_k\boldsymbol{p}^k,\\ \lambda_k=-\dfrac{\nabla f(\boldsymbol{x}^k)^{\mathrm{T}}\boldsymbol{p}^k}{(\boldsymbol{p}^k)^{\mathrm{T}}H(\boldsymbol{x}^k)\boldsymbol{p}^k},\\ \boldsymbol{p}^{k+1}=-\nabla f(\boldsymbol{x}^{k+1})+\beta_k\boldsymbol{p}^k,\\ \beta_k=\dfrac{\nabla f(\boldsymbol{x}^{k+1})^{\mathrm{T}}H(\boldsymbol{x}^k)\boldsymbol{p}^k}{(\boldsymbol{p}^k)^{\mathrm{T}}H(\boldsymbol{x}^k)\boldsymbol{p}^k}.\end{cases}$$

当迭代次数 $k\leqslant n$,已达到精度的要求,$\boldsymbol{x}^k$就是要求的近似最优解. 否则,可将前 n 步作为第一个循环,同时以所得的 $\boldsymbol{x}^n$作为新的初始点重新开始,进行第二个循环,重复进行,直至满足要求的精度为止.

牛顿方向(Newton direction)　求解无约束最优化问题中的一个概念. 指向二次函数最优点的方向. 对于 $\mathbf{R}^n$ 上的非二次函数 f,设 f 在 $\mathbf{R}^n$ 中二次连续可微,点 $\boldsymbol{x}^k\subset\mathbf{R}^n$,$\nabla f(\boldsymbol{x}^k)$是 f 在 $\boldsymbol{x}^k$处的梯度,$H(\boldsymbol{x}^k)$是 f 在 $\boldsymbol{x}^k$处的黑塞矩阵,则 f 在 $\boldsymbol{x}^k$处的牛顿方向是$-H(\boldsymbol{x}^k)^{-1}\nabla f(\boldsymbol{x}^k)$.

修正牛顿法(modified Newton method)　寻求无约束最优化问题极小点的方法. 按目标函数在迭代点处的牛顿方向,进行一维搜索迭代,以寻求无约束最优化问题的极小点的方法. 设 f 是目标函数,$\boldsymbol{x}^k$是当前迭代点,其迭代公式为

$$\begin{cases}\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\lambda_k\boldsymbol{p}^k,\\ \boldsymbol{p}^k=-H(\boldsymbol{x}^k)^{-1}\nabla f(\boldsymbol{x}^k),\\ \lambda_k:\ \min\limits_{\lambda} f(\boldsymbol{x}^k+\lambda\boldsymbol{p}^k).\end{cases}$$

修正牛顿法的收敛速度很快,当 f 的二阶导数及其黑塞矩阵的逆阵便于计算时,使用这种方法非常有效. 修正牛顿法有时也简称牛顿法. 为了区别于不进行一维搜索的古典牛顿法,这里加了"修正"二字.

拟牛顿条件(quasi-Newton condition)　搜索方向中迭代矩阵的条件. 指搜索方向中的迭代矩阵具有近似于目标函数的黑塞矩阵的逆矩阵特性的条件. 设 f 是 $\mathbf{R}^n$ 上的非二次函数,$\boldsymbol{x}^k,\boldsymbol{x}^{k+1}\in\mathbf{R}^n$ 是前后两迭代点,H^{k+1}是第 $k+1$ 次迭代矩阵,则 H^{k+1}要满足的拟牛顿条件是

$$\boldsymbol{x}^{k+1}-\boldsymbol{x}^k=H^{k+1}[\nabla f(\boldsymbol{x}^{k+1})-\nabla f(\boldsymbol{x}^k)].$$

拟牛顿法(quasi-Newton method)　寻求无约束最优化问题的一类方法. 在修正牛顿法的迭代公式中,构造满足拟牛顿条件的对称正定矩阵 H^k 去逼近目标函数黑塞矩阵的逆阵$H(\boldsymbol{x}^k)^{-1}$,以建立搜索方向的方法. 这样,拟牛顿法不必计算修正牛顿法迭代公式中的二阶导数矩阵$H(\boldsymbol{x}^k)$,也不必计算其逆矩阵$H(\boldsymbol{x}^k)^{-1}$. 称这样修改的牛顿法为拟牛顿法.

尺度矩阵(metric matrix)　拟牛顿法修正公式中的矩阵. 第 $k+1$ 次迭代尺度矩阵即

$$H^{k+1}=H^k+\frac{\Delta\boldsymbol{x}^k(\Delta\boldsymbol{x}^k)^{\mathrm{T}}}{(\Delta\boldsymbol{x}^k)^{\mathrm{T}}\Delta g^k}-\frac{H^k\Delta g^k(\Delta g^k)^{\mathrm{T}}H^k}{(\Delta g^k)^{\mathrm{T}}H^k\Delta g^k},$$

式中 $\Delta g^k=\nabla f(\boldsymbol{x}^{k+1})-\nabla f(\boldsymbol{x}^k)$,$\Delta\boldsymbol{x}^k=\boldsymbol{x}^{k+1}-\boldsymbol{x}^k$.

校正矩阵(correction matrix)　拟牛顿法中的修正矩阵. 即矩阵

$$\Delta H^k=\frac{\Delta\boldsymbol{x}^k(\Delta\boldsymbol{x}^k)^{\mathrm{T}}}{(\Delta\boldsymbol{x}^k)^{\mathrm{T}}\Delta g^k}-\frac{H^k\Delta g^k(\Delta g^k)^{\mathrm{T}}H^k}{(\Delta g^k)^{\mathrm{T}}H^k\Delta g^k},$$

式中 $\Delta g^k=\nabla f(\boldsymbol{x}^{k+1})-\nabla f(\boldsymbol{x}^k)$,$\Delta\boldsymbol{x}^k=\boldsymbol{x}^{k+1}-\boldsymbol{x}^k$,$H^k$是尺度矩阵(参见"拟牛顿法").

变尺度法(variable metric method)　一种拟牛顿法. 指用尺度矩阵作为拟牛顿法中的对称正定迭代矩阵的方法. 此法首先由戴维顿(Davidon,W. C.)于 1959 年提出,后经弗莱彻(Fletcher,R.)和鲍威尔(Powell,M. J. D.)改进,故亦称 DFP 法,或称 DFP 变尺度法.

DFP 法(DFP method)　即"变尺度法".

BFGS 矩阵(BFGS matrix)　尺度矩阵的推广. 其第 $k+1$ 次迭代矩阵为

$$H^{k+1}=H^k+\frac{\Delta\boldsymbol{x}^k(\Delta\boldsymbol{x}^k)^{\mathrm{T}}}{(\Delta\boldsymbol{x}^k)^{\mathrm{T}}\Delta g^k}\left[1+\frac{(\Delta g^k)^{\mathrm{T}}H^k\Delta g^k}{(\Delta\boldsymbol{x}^k)^{\mathrm{T}}\Delta g^k}\right]-\frac{1}{(\Delta\boldsymbol{x}^k)^{\mathrm{T}}\Delta g^k}[\Delta\boldsymbol{x}^k(\Delta g^k)^{\mathrm{T}}H^k+H^k\Delta g^k(\Delta\boldsymbol{x}^k)^{\mathrm{T}}].$$

式中 $\Delta g^k=\nabla f(\boldsymbol{x}^{k+1})-\nabla f(\boldsymbol{x}^k)$,$\Delta\boldsymbol{x}^k=\boldsymbol{x}^{k+1}-\boldsymbol{x}^k$.

BFGS 法(BFGS method)　一种拟牛顿法. 指用 BFGS 矩阵作为拟牛顿法中的对称正定迭代矩阵的方法. 此法是 1970 年前后由柏萝登(Broyden,C. G.)、弗莱彻(Fletcher,R.)、戈德福布(Goldfarb,D.),以及生纳(Shanno,D. F.)所研究,故得名. 由于 BFGS 法对一维搜索的精度要求不高,并且由迭代产生的 BFGS 矩阵不易变为奇异矩阵,因而 BFGS 法比 DFP 法在计算中具有更好的数值稳定性.

坐标轮换法(univariate search technique)　亦称变量轮换法. 一种求无约束最优化问题的降维方法. 其迭代过程是沿不同的坐标方向轮换地进行搜索. 设初始点为 $\boldsymbol{x}^0$,先只改变一个变量,其他变量视为常数,进行一维搜索,得到最优点 $\boldsymbol{x}^1$;再换一个变量,同样进行一维搜索,得到最优点 $\boldsymbol{x}^2$;如此继续下去,直到完成一个循环,即对 n 个变量都轮换完毕,得到最优点 $\boldsymbol{x}^n$. 若 $\|\boldsymbol{x}^n-\boldsymbol{x}^0\|\leqslant\varepsilon$($\boldsymbol{x}^0$为这一循环的初始点,$\varepsilon$ 是精度要求),则得近似最优解 $\boldsymbol{x}^*=\boldsymbol{x}^n$. 否则,令 $\boldsymbol{x}^0=\boldsymbol{x}^n$,再重复以上的迭代过程.

变量轮换法(univariate search technique)　即"坐标轮换法".

步长加速法(step length acceleration method)

亦称模式搜索法或模矢法. 一种追循谷线(脊线)方向加速寻优的方法. 它既简单又易于编制计算机程序. 方法由两种"移动"组成:一是探测性移动,即沿每一个固定的坐标方向作一定步长的探索,探索出一个好方向;另一是模式性移动,即在一轮探测性移动完成之后,沿探索所得的好方向进行加速移动. 反复进行,直至得到满足精度要求为止. 此法由胡克(Hooke,R.)和基夫斯(Jeeves,T. A.)于 1961 年提出.

模式搜索法(pattern search method) 即"步长加速法".

方向加速法(direction acceleration method) 在坐标轮换法的基础上加以改进的方法. 指在坐标轮换法中,一维搜索是沿坐标方向进行的,收敛速度较慢,故考虑从方向上作改进,以产生加速的方向. 1964 年,鲍威尔(Powell,M. J. D.)提出了这个方法. 它以二次正定函数为背景,以共轭方向为基础,是目前求无约束最优化问题直接法中最有效的方法之一.

单纯形调优法(simplex evolutionary method) 一种多维直接搜索法. 指先给定多维空间中一个初始单纯形(参见"单纯形"),求出单纯形各个顶点的目标函数值,并加以比较,丢掉其中最坏的点,代之以新点,从而构成一新的单纯形. 如此迭代下去,逐步逼近最优点. 此法于 1962 年由斯盆得莱(Spendly,W.)、赫斯蒂(Hext,G. R.)、海姆斯瓦尔斯(Himsworth,F. R.)等人提出,1965 年涅尔得(Nelder,J. A.)和梅得(Mead,R.)两人加以改进.

单纯形加速(simplex acceleration) 单纯形法的推广. 指单纯形法中的扩展、压缩、缩边. 例如,设 $f(\boldsymbol{x})=f(x_1,x_2)$ 为二维无约束极小化问题的目标函数,在 x_1x_2 平面上取不在同一直线上的三个点 $\boldsymbol{x}^1$, $\boldsymbol{x}^2$ 和 $\boldsymbol{x}^3$,并以它们为顶点,构成一单纯形(即三角形). 算出各顶点的函数值 $f(\boldsymbol{x}^1)$, $f(\boldsymbol{x}^2)$ 和 $f(\boldsymbol{x}^3)$. 如果 $f(\boldsymbol{x}^1)>f(\boldsymbol{x}^2)>f(\boldsymbol{x}^3)$,则 $\boldsymbol{x}^1$ 点最差,$\boldsymbol{x}^3$ 点最好. 为了寻找极小点,应向最差点的反对称方向进行搜索. 设线段 $\boldsymbol{x}^2\boldsymbol{x}^3$ 的中点为 $\boldsymbol{x}^4$,在 $\boldsymbol{x}^1\boldsymbol{x}^4$ 的延长线上取点 $\boldsymbol{x}^5$,使 $\boldsymbol{x}^5=\boldsymbol{x}^4+(\boldsymbol{x}^4-\boldsymbol{x}^1)=2\boldsymbol{x}^4-\boldsymbol{x}^1$,$\boldsymbol{x}^5$ 称为 $\boldsymbol{x}^1$ 关于 $\boldsymbol{x}^4$ 的反射点. 算出 $\boldsymbol{x}^5$ 的函数值 $f(\boldsymbol{x}^5)$.

1. 扩展. 若 $f(\boldsymbol{x}^5)<f(\boldsymbol{x}^3)$,说明搜索方向有效,继续沿 $\boldsymbol{x}^1\boldsymbol{x}^5$ 向前扩展,取 $\boldsymbol{x}^6=\boldsymbol{x}^4+\alpha(\boldsymbol{x}^4-\boldsymbol{x}^1)$,其中 α 为扩展因子.

2. 压缩. 若 $f(\boldsymbol{x}^2)<f(\boldsymbol{x}^5)<f(\boldsymbol{x}^1)$,说明 $\boldsymbol{x}^5$ 点走得太远,应压缩,取 $\boldsymbol{x}^7=\boldsymbol{x}^4+\beta(\boldsymbol{x}^5-\boldsymbol{x}^4)$,其中 β 为压缩因子.

3. 缩边. 若 $\boldsymbol{x}^1\boldsymbol{x}^4$ 方向上所有点的函数值都大于 $f(\boldsymbol{x}^1)$,则不能沿此方向搜索. 这时,以 $\boldsymbol{x}^3$ 为中心进行缩边. 例如,使顶点 $\boldsymbol{x}^1$ 和 $\boldsymbol{x}^2$ 向 $\boldsymbol{x}^3$ 移近一半距离,得新单纯形. 以此为基础再进行寻优.

随机跳跃法(random jump method) 利用随机数寻求最优点的一种搜索方法. 设要求 $f(\boldsymbol{x})$ 的极小点,$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathrm{R}^n$, $l_i\leqslant x_i\leqslant u_i\ (i=1,2,\cdots,n)$,式中 l_i 和 u_i 为变量 $x_i\ (i=1,2,\cdots,n)$ 的上限和下限. 再设 n 个随机数 $r_1,r_2,\cdots,r_n$ 在 0 和 1 之间均匀分布. 应用每组随机数求 $\boldsymbol{x}$:$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}=(l_1+r_1(u_1-l_1),l_2+r_2(u_2-l_2),\cdots,l_n+r_n(u_n-l_n))^{\mathrm{T}}$,并计算 $f(\boldsymbol{x})$,取使 $f(\boldsymbol{x})$ 最小的 $\boldsymbol{x}$,它就作为所求的极小点.

随机走步法(random walk method) 利用随机数确定搜索方向的方法. 为求 $f(\boldsymbol{x})$ 的极小点,设已产生点列 $\boldsymbol{x}^1,\boldsymbol{x}^2,\cdots,\boldsymbol{x}^k$,第 $k+1$ 步得到的点由第 k 步得到的点修正产生:$\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\lambda\boldsymbol{u}^k$,$\lambda$ 为步长,$\boldsymbol{u}^k$ 为第 k 步产生的单位随机向量,单位随机向量的定义为

$$\boldsymbol{u}=\frac{(r_1,r_2,\cdots,r_n)^{\mathrm{T}}}{\sqrt{r_1^2+r_2^2+\cdots+r_n^2}},$$

$\|\boldsymbol{u}\|=1$,式中 $r_1,r_2,\cdots,r_n$ 为区间 $[-1,1]$ 内的一组随机数. 若不采用固定步长,设沿某一单位随机方向 $\boldsymbol{u}^k$ 能使目标函数值减少,则或者沿该方向加大步长进行搜索,或者求最优步长,令

$$f(\boldsymbol{x}^k+\lambda_k^*\boldsymbol{u}^k)=\min_{\lambda_k} f(\boldsymbol{x}^k+\lambda_k\boldsymbol{u}^k),$$

得 $\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\lambda_k^*\boldsymbol{u}^k$.

约束最优化问题(constrained optimization problem) 具有约束条件的非线性规划问题. 极小化问题的一般形式为

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t.}\ g_i(\boldsymbol{x})\leqslant 0\quad (i=1,2,\cdots,p),\\ \qquad h_j(\boldsymbol{x})=0\quad (j=1,2,\cdots,q).\end{cases}$$

仅有等式约束条件的约束最优化问题,可采用消元法、拉格朗日乘子法或罚函数法,将其化为无约束最优化问题求解. 对于含有等式约束和不等式约束条件的最优化问题,可采用以下方法:将不等式约束化为等式约束;将约束问题化为无约束问题;将非线性规划问题用线性逼近的方法来近似求解;在可行域中沿某方向作一维搜索,寻求最优解.

可行方向(feasible direction) 求解约束最优化问题中,保持可行性的搜索方向. 设 $f(\boldsymbol{x})$, $g_i(\boldsymbol{x})$ 和 $h_j(\boldsymbol{x})$(参见"约束最优化问题")都具有一阶连续偏导数,$X=\{\boldsymbol{x}\mid g_i(\boldsymbol{x})\leqslant 0(i=1,2,\cdots,p);h_j(\boldsymbol{x})=0(j=1,2,\cdots,q)\}$. 设 $\boldsymbol{x}^0$ 是一个可行点,即 $\boldsymbol{x}^0\in X$. 对于某一方向 $\boldsymbol{p}$ 来说,若存在数 $\lambda_0>0$,使对于任意的 $\lambda(0\leqslant\lambda\leqslant\lambda_0)$ 均有 $\boldsymbol{x}^0+\lambda\boldsymbol{p}\in X$ 成立,则 $\boldsymbol{p}$ 是点 $\boldsymbol{x}^0$ 处的一个可行方向.

可行下降方向(feasible descent direction) 使目标函数的值下降的可行方向(参见“可行方向”). 如果可行点 $\tilde{x}$ 不是问题的极小点,就需要在点 $\tilde{x}$ 处寻找能使目标函数 $f(x)$ 的值下降的可行方向 p,它为 $f(x)$ 在点 $\tilde{x}$ 处的可行下降方向,即满足 $\nabla f(\tilde{x})^{\mathrm{T}}p<0$ 的方向 p.

起作用约束(active constraint) 亦称紧约束或积极约束. 在某可行点处使不等式约束成为等式的约束. 对于可行点 $\tilde{x}$,当 $g_i(\tilde{x})=0$ 时点 $\tilde{x}$ 处于这个约束形成的可行域的边界上,起到了限制作用,故称这一约束为点 $\tilde{x}$ 处的起作用约束. 等式约束对所有可行点都是起作用约束.

可行方向法(method of feasible directions) 利用可行下降方向寻求约束非线性规划问题极小解的方法. 考虑约束非线性规划问题

$$\min_{x\in X} f(x),$$

其中 $X=\{x\in\mathrm{R}^n \mid g_i(x)\leqslant 0, i=1,2,\cdots,p\}$. 设目标函数 $f(x)$ 和约束函数 $g_i(x)$ $(i=1,2,\cdots,p)$ 具有一阶连续偏导数. 从某一点 $x^k\in X$ 出发,首先找出该点处的可行下降方向 p^k,即它要同时满足 $\nabla f(x^k)^{\mathrm{T}}p^k<0$ 和 $\nabla g_i(x^k)^{\mathrm{T}}p^k>0$,然后确定在这方向上移动的步长 λ_k,得到更好的近似点 x^{k+1}: $x^{k+1}=x^k+\lambda_k p^k\in X$. 若点 x^{k+1} 满足精度要求,则迭代停止. 否则,再在点 x^{k+1} 处找出可行下降方向,继续迭代,直至满足精度要求为止. 最早的可行方向法是藻腾狄克(Zoutendijk,G.)于 1960 年提出的.

近似线性化法(approximation linearized method) 一种求解约束非线性规划问题的方法. 对于具有线性约束条件的约束非线性规划问题,通过不断对目标函数作近似线性化处理,利用求线性规划的方法进行求解的方法. 考虑约束非线性规划问题

$$\begin{cases}\min f(x),\\ \text{s.t. } Ax\geqslant \mathbf{0},\\ \qquad x\geqslant \mathbf{0},\end{cases}$$

记 $X=\{x\in R^n \mid Ax\geqslant b, x\geqslant \mathbf{0}\}$). 设 f 在 X 上具有一阶连续偏导数. 取任一点 $x^0\in X$,则有 $f(x)\approx f(x^0)+\nabla f(x^0)^{\mathrm{T}}(x-x^0)$. 求线性规划问题

$$\min_{x\in X}[f(x^0)+\nabla f(x^0)^{\mathrm{T}}(x-x^0)]$$

的最优解,等价于求线性规划问题 $\min_{x\in X}[\nabla f(x^0)^{\mathrm{T}}x]$ 的最优解 $\bar{x}$.

1. 若 $\nabla f(x^0)^{\mathrm{T}}(\bar{x}-x^0)=0$,则 $\bar{x}$ 为问题的最优解,停止迭代.

2. 若 $\nabla f(x^0)^{\mathrm{T}}(\bar{x}-x^0)\neq 0$,则进行一维搜索

$$\min_{\lambda\in[0,1]} f(x^0+\lambda(\bar{x}-x^0)),$$

其最优解 λ_0 满足 $0<\lambda_0<1$. 令 $x^1=x^0+\lambda_0(\bar{x}-x^0)$, x^1 是 x^0 与 $\bar{x}$ 的连线上使目标函数 $f(x)$ 取最优值的点,$x^1\in X$,且 $f(x^1)<f(x^0)$. 得到 x^1 后,再在 x^1 处对 $f(x)$ 进行线性化逼近,重复上述步骤. 若在点 x^k 发生第一种情况,则在此点满足库恩-塔克尔条件(参见“库恩-塔克尔条件”);若第一种情况总不发生,则得到有界点列 $\{x^k\}$,其极限点满足库恩-塔克尔条件. 此外,当 $f(x)$ 为凸函数时,它收敛于

$$\min_{x\in X}[\nabla f(x^0)^{\mathrm{T}}x]$$

的最优解.

此法由弗兰克(Frank,P. H.)和沃尔夫(Wolfe,P.)于 1956 年提出,故也称为 F-W 法. 此外,还有藻腾狄克(Zoutendijk,G.)于 1960 年提出的近似线性化方法. 同时,也可将近似线性化推广到对非线性约束函数的处理.

F-W 法(F-W method) 即“近似线性化法”.

可接受步长法(acceptable steplength method) 以可接受的步长取代最优步长的一维搜索方法. 比起寻求最优步长,它可以大大减少计算量. 设 $f(x)$ 为定义在 R^n 中包含约束集 X 的某一开集上的实值函数,且具有二阶连续偏导数. $x^k,\bar{x}^k\in X$,在 $[0,1]$ 中求出使

$$f(x^k+\lambda(\bar{x}^k-x^k)) \leqslant f(x^k)+(\lambda/2)\nabla f\cdot(x^k)(\bar{x}^k-x^k)$$

成立的最大的 λ,它即为可接受步长.

正则点(regular point) 与约束条件有关的可行点. 在约束非线性规划问题(NP)中,设 $\bar{x}$ 是满足约束条件的点,I 是使得 $g_i(\bar{x})=0$ 的那些下标 i 的集合(参见“非线性规划”). 若梯度向量 $\nabla g_i(\bar{x})$ $(i\in I)$, $\nabla h_j(\bar{x})$ $(j=1,2,\cdots,q)$ 是线性无关的,则称 $\bar{x}$ 为约束条件的一个正则点.

库恩-塔克尔条件(Kuhn-Tucker condition) 判定约束非线性规划问题的某可行点为极小点的必要条件. 对于凸规划来说,则是判别极小点的充分必要条件. 对于约束非线性规划问题(NP)(参见“非线性规划”),设其中 $f(x)$, $g_i(x)$ $(i=1,2,\cdots,p)$ 和 $h_j(x)$ $(j=1,2,\cdots,q)$ 在 R^n 的某一开集上一阶连续可微,x^* 是问题的极小点,且是约束条件的正则点,则存在向量 $\lambda^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_p^*)^{\mathrm{T}}$ 及 $\mu^*=(\mu_1^*,\mu_2^*,\cdots,\mu_q^*)^{\mathrm{T}}$,使得

$$\begin{cases}\nabla f(x^*)+\sum_{i=1}^{p}\lambda_i^*\nabla g_i(x^*)\\ \qquad +\sum_{j=1}^{q}\mu_j^*\nabla h_j(x^*)=0,\\ \lambda_i^* g_i(x^*)=0 \quad (i=1,2,\cdots,p),\\ \mu_j^*\geqslant 0 \qquad\quad (j=1,2,\cdots,q).\end{cases}$$

此即为所考虑约束非线性规划问题(NP)的库恩-塔克尔条件,也称一阶必要条件. λ_i^* $(i=1,2,\cdots,p)$ 及

μ_j^* ($j=1,2,\cdots,q$)称为库恩-塔克尔乘子. 由上述库恩-塔克尔条件可知,只有当$g_i(\boldsymbol{x})$在 $\boldsymbol{x}^*$ 点为起作用约束时,可以有 $\lambda_i^* \neq 0$;否则,$\lambda_i^*=0$. 1951 年,库恩(Kuhn,H. W.)和塔克尔(Tucker,A. W.)证明了这一条件,为非线性规划奠定了重要理论基础.

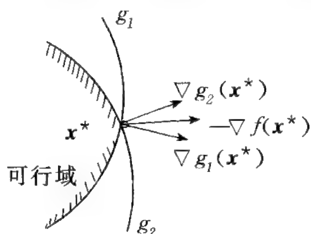
一阶必要条件(first order necessary conditions) 即"库恩-塔克尔条件".

库恩-塔克尔乘子(Kuhn-Tucker multiplier) 见"库恩-塔克尔条件".

库恩-塔克尔条件的几何解释(geometric interpretation of Kuhn-Tucker condition) 一种非线性规划理论中的基本条件的几何表述. 对于非线性规划问题(NP)(参见"非线性规划"),设 $p=2,q=0$, $\boldsymbol{x}^*$ 是(NP)的极小点. 由库恩-塔克尔条件可知, $-\nabla f(\boldsymbol{x}^*)$可由 $\nabla g_1(\boldsymbol{x}^*)$和 $\nabla g_2(\boldsymbol{x}^*)$的非负线性组合表示,即

$$-\nabla f(\boldsymbol{x}^*) = \lambda_1^* \nabla g_1(\boldsymbol{x}^*) + \lambda_2^* \nabla g_2(\boldsymbol{x}^*).$$

目标函数 f 在某点 $\boldsymbol{x}$ 处的负梯度方向$-\nabla f(\boldsymbol{x})$是 f 在该点处的最速下降方向. 若继续搜索的方向与$-\nabla f(\boldsymbol{x})$的夹角小于 90°,则 $f(\boldsymbol{x})$还能下降. 梯度 $\nabla g_i(\boldsymbol{x})$ ($i=1,2$)的方向是相应约束函数 $g_i(\boldsymbol{x})$的值增加最快的方向,它垂直于该约束函数的等值面(等值线)在相应点处的切平面(切线). 在点 $\boldsymbol{x}^*$ 处,由于$-\nabla f(\boldsymbol{x}^*)$处于 $\nabla g_1(\boldsymbol{x}^*)$与 $\nabla g_2(\boldsymbol{x}^*)$的夹角范围内,故已找不到使目标函数值进一步下降的可行方向,从而 $\boldsymbol{x}^*$ 可能是最优点(见图). 反之,若某点的$-\nabla f(\boldsymbol{x})$处于其 $\nabla g_1(\boldsymbol{x})$与 $\nabla g_2(\boldsymbol{x})$的夹角之外,则还能在可行域内找到更好的点,因而它不可能是最优点.



库恩-塔克尔充分条件(Kuhn-Tucker sufficient condition) 判定约束非线性规划问题某可行点为极小点的充分条件. 对于凸规划来说,则是判别极小点的充分必要条件. 对于有约束非线性规划问题(NP)(参见"非线性规划"),设 $f(\boldsymbol{x}),g_i(\boldsymbol{x})$($i=1,2,\cdots,p$)和 $h_j(\boldsymbol{x})$($j=1,2,\cdots,q$)二阶连续可微,且 $\boldsymbol{x}^*$ 是可行点. 如果存在向量 $\boldsymbol{\lambda}^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_p^*)^{\mathrm{T}}$ 和 $\boldsymbol{\mu}^*=(\mu_1^*,\mu_2^*,\cdots,\mu_q^*)^{\mathrm{T}}$ 满足库恩-塔克尔条件,且对满足条件

$$\begin{cases} z^{\mathrm{T}}\nabla g_i(\boldsymbol{x}^*)=0 & (i\in I,\lambda_i^*>0),\\ z^{\mathrm{T}}\nabla g_i(\boldsymbol{x}^*)\leqslant 0 & (i\in I,\lambda_i^*=0),\\ z^{\mathrm{T}}\nabla h_j(\boldsymbol{x}^*)=0 & (j=1,2,\cdots,q)\end{cases}$$

的任意非零向量 z,有

$$\boldsymbol{z}^{\mathrm{T}}\Big[\nabla^2 f(\boldsymbol{x}^*)+\sum_{i=1}^{p}\lambda_i^*\nabla^2 g_i(\boldsymbol{x}^*) + \sum_{j=1}^{q}\mu_j^*\nabla^2 h_j(\boldsymbol{x}^*)\Big]\boldsymbol{z}>0,$$

式中 I 是使 $g_i(\boldsymbol{x}^*)=0$ 的下标 i 的集合,$\nabla^2 f(\boldsymbol{x}^*)$, $\nabla^2 g_i(\boldsymbol{x}^*)$($i=1,2,\cdots,p$)和$\nabla^2 h_j(\boldsymbol{x}^*)$($j=1,2,\cdots,q$)依次是 $f(\boldsymbol{x}),g_i(\boldsymbol{x})$和 $h_j(\boldsymbol{x})$在点 $\boldsymbol{x}^*$ 处的黑塞矩阵. 则上述条件为库恩-塔克尔充分条件.

二阶充分条件(second order sufficient conditions) 即"库恩-塔克尔充分条件".

制约函数法(constrained function method) 亦称序列无约束极小化技术. 是将求解约束非线性规划问题转化为求解一系列无约束最优化问题的方法. 考虑约束非线性规划问题(NP)

$$\begin{cases}\min f(\boldsymbol{x}),\\ \text{s.t. } \boldsymbol{x}\in X,\end{cases}$$

其中 $X=\{\boldsymbol{x}\in \mathrm{R}^n \mid g_i(\boldsymbol{x})\leqslant 0, i=1,2,\cdots,p; h_j(\boldsymbol{x})=0, j=1,2,\cdots,q\}$是(NP)的可行域. 将有约束的问题(NP)转化为无约束极小化问题求解的原始想法是:设法适当地加大不可行点处的目标函数值,使不可行点不能成为相应无约束极小化问题的最优解. 具体地说,是预先选定一个很大的数 M,并对可行域 X 构造如下的制约函数:

$$P(\boldsymbol{x})=\begin{cases}0 & (\boldsymbol{x}\in X);\\ M & (\boldsymbol{x}\notin X).\end{cases}$$

然后,利用 $P(\boldsymbol{x})$作一个(NP)的增广目标函数 $F(\boldsymbol{x})=f(\boldsymbol{x})+P(\boldsymbol{x})$. 由于在可行点处 $F(\boldsymbol{x})$的值与$f(\boldsymbol{x})$的值相同,而在不可行点处对应的 $F(\boldsymbol{x})$值很大,所以,相应的以增广目标函数为目标函数的无约束极小化问题 $\min F(\boldsymbol{x})$ 的最优解,必定也是约束非线性规划问题(NP)的最优解. 这类利用制约函数将求解问题(NP)转化为求解无约束极小化问题的方法称为制约函数法. 制约函数有惩罚函数和障碍函数两种. 对应于这两种函数有求解约束非线性规划问题的外点法和内点法. 此外,还有序列加权因子法.

序列无约束极小化技术(sequential unconstrained minimization technique) 简称 SUMT. 见"制约函数法".

惩罚函数(penalty function) 亦称处罚函数. 一类制约函数. 对于约束非线性规划问题

$$\begin{cases}\min f(\boldsymbol{x}),\\ \text{s.t. } g_i(\boldsymbol{x})\leqslant 0 & (i=1,2,\cdots,p),\\ \qquad h_j(\boldsymbol{x})=0 & (j=1,2,\cdots,q),\end{cases}$$

它的形如

$$P(\boldsymbol{x},M)=M\sum_{i=1}^{p}[\max(0,g_i(\boldsymbol{x}))]^2 + \frac{M}{2}\sum_{j=1}^{q}[h_j(\boldsymbol{x})]^2 \quad (M>0)$$

的制约函数称为惩罚函数，其中的 M 叫惩罚因子(或罚参数).

惩罚因子(penalty factor) 见“惩罚函数”.

外点法(exterior point method) 亦称惩罚函数法或外惩罚函数法. 将求解约束非线性规划问题转化为求解一系列无约束极小化问题的一类制约函数法(参见“制约函数法”). 对于约束非线性规划问题(NP)

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t.}\ g_i(\boldsymbol{x})\leqslant 0\quad (i=1,2,\cdots,p),\\ \qquad h_j(\boldsymbol{x})=0\quad (j=1,2,\cdots,q),\end{cases}$$

求解过程如下：选取一系列递增的惩罚因子 $M_k(k=1,2,\cdots)$，作相应惩罚函数列：

$$P(\boldsymbol{x},M_k)=M_k\sum_{i=1}^{p}[\max(0,g_i(\boldsymbol{x}))]^2+\frac{M_k}{2}\sum_{j=1}^{q}[h_j(\boldsymbol{x})]^2\quad (k=1,2,\cdots),$$

将求解问题(NP)转化为求解一系列无约束极小化问题

$$\min\ \{f(\boldsymbol{x})+P(\boldsymbol{x},M_k)\}\quad (k=1,2,\cdots).\quad (P_k)$$

设 $\boldsymbol{x}^k$ 是(P_k)的最优解，则在一定条件下可以证明点列$\{\boldsymbol{x}^k\}$存在极限点 $\boldsymbol{x}^*$，并且 $\boldsymbol{x}^*$ 是问题(NP)的最优解. 用此法求解(NP)是将“惩罚”加于不可行点，通过逐步增大惩罚因子 M_k，迫使解的迭代点列$\{\boldsymbol{x}^k\}$从可行域外部向位于可行域边界上的某最优解接近，所以称为外点法. 由于利用了惩罚函数进行求解，故也称惩罚函数法或外惩罚函数法.

外点法的经济解释是：一个采购者的经济决策问题，即如果把目标函数视为采购者决策时要花费的总“价格”数，约束条件视为某种“规定”，采购者在“规定”范围内买东西不罚款；若违反“规定”，则按罚款政策罚款. 因此，采购者付出的总代价应是总的“价格”数与总的罚款数之和. 采购者要以总代价作为最终目标. 当把罚款政策定得非常苛刻时，即违反“规定”，则重罚，迫使采购者做决策时不敢违反“规定”，否则，他的总代价就不可能省. 在数学上则表现在惩罚因子 M_k 足够大时，无约束极小化问题的最优解 $\boldsymbol{x}^*$ 应该满足诸约束条件，因而是约束极小化问题的最优解.

外惩罚函数法(external penalty function method) 即“外点法”.

障碍函数(barrier function) 亦称内惩罚函数、围墙函数或碰壁函数. 一类制约函数. 对于约束非线性规划问题

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t.}\ g_i(\boldsymbol{x})\leqslant 0\quad (i=1,2,\cdots,p),\end{cases}$$

它的形如

$$B(\boldsymbol{x},r)=-r\sum_{i=1}^{p}\frac{1}{g_i(\boldsymbol{x})}\quad (r>0)$$

或

$$B(\boldsymbol{x},r)=-r\sum_{i=1}^{l}\lg[-g_i(\boldsymbol{x})]\quad (r>0)$$

的制约函数称为障碍函数，其中的 r 称为障碍因子(或内罚参数).

障碍因子(barrier factors) 见“障碍函数”.

内点法(interior point method) 亦称障碍函数法或内惩罚函数法. 求解带不等式约束的非线性规划问题的一类制约函数法(参见“制约函数法”). 对于带不等式约束条件的非线性规划问题，从可行域内部的某一点 $\boldsymbol{x}^0$出发，利用障碍函数构成增广目标函数，将问题转化为求解一系列无约束极小化问题的方法. 随着障碍因子 r_k 的逐步减小，即 $r_1>r_2>\cdots>r_k>\cdots>0$，障碍函数所起的作用也越来越小，因而，求出的一系列最优解 $\boldsymbol{x}^k(k=1,2,\cdots)$也逐步逼近原约束非线性规划问题的最优解. 若原问题的最优解在可行域的边界上，则随着 r_k 的减小，障碍作用逐步降低，所求出的最优解序列不断靠近边界，直至满足某一精度要求为止. 由于此法利用了障碍函数进行求解，故也称障碍函数法或内惩罚函数法.

障碍函数法(barrier function method) 即“内点法”.

初始内点的求法(method for finding initial interior point) 求带不等式约束非线性规划问题的可行域内部点的方法. 对带不等式约束的非线性规划问题

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t.}\ g_i(\boldsymbol{x})\leqslant 0\quad (i=1,2,\cdots,p),\end{cases}$$

先任找一点 $\boldsymbol{x}^0$为初始点，令 $S_0=\{i\,|\,g_i(\boldsymbol{x}^0)\geqslant 0,1\leqslant i\leqslant p\}$，$T_0=\{i\,|\,g_i(\boldsymbol{x}^0)<0,1\leqslant i\leqslant p\}$. 若 S_0 为空集，则 $\boldsymbol{x}^0$为初始内点；若 S_0 非空，则以 S_0 中的约束函数为假拟目标函数，并以 T_0 中的约束函数为障碍函数，构成一无约束极小化问题. 求解这一问题，可得一个新点 $\boldsymbol{x}^1$. 然后检验 $\boldsymbol{x}^1$，若仍不是内点，则继续进行，并减小障碍因子 r，直到求出一个内点为止.

序列加权因子法(sequential weighted factor technique method) 求解约束非线性规划问题的方法. 对于约束非线性规划问题

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t.}\ g_i(\boldsymbol{x})\leqslant 0\quad (i=1,2,\cdots,p),\\ \qquad h_j(\boldsymbol{x})=0\quad (j=1,2,\cdots,q),\end{cases}$$

用单纯形调优法与惩罚函数相结合进行迭代求解. 每一轮迭代用单纯形调优法(参见“单纯形调优法”)求无约束极小化问题，而在惩罚项上的加权(惩罚因子)初始时取 1，以后由前次迭代结果给出.

锯齿算法(sawtooth algorithm)　亦称花边算法.沿约束边界寻优的一种方法.是梯度法应用于约束非线性规划问题的情形.先在可行域内沿目标函数负梯度方向前进,当到达可行域外时,再沿约束函数的梯度方向折回来,搜索路线是沿着约束边界曲折前进的,在目标函数负梯度和约束函数梯度之间不断变换方向.

花边算法(lace algorithm)　即"锯齿算法".

梯度投影法(gradient projection method)　利用梯度的投影技巧求约束非线性规划问题最优解的一种方法.求带线性约束的非线性规划问题更为有效.它是从一个基本可行解开始,由约束条件确定出凸约束集边界上梯度的投影,以便求出下次的搜索方向和步长.每次搜索后,都要进行检验,直到满足精度要求为止.这种方法是罗森(Rosen,J.B.)于1960年提出的,戈德福布(Goldfarb,D.)和拉匹塔斯(Lapidus,L.)于1968年作了改进.

广义乘子法(generalized method of multipliers)　一种求解约束非线性规划问题的方法.即把古典的拉格朗日乘子法与外点法结合起来求最优解的方法.在惩罚因子 M 适当大的情况下,借助于调节拉格朗日乘子 $\boldsymbol{\lambda}$ 来逐次逼近非线性规划问题的最优解.这种方法是由赫斯泰尼斯(Hestenes,M.R.)和鲍威尔(Powell,M.J.D.)于1969年提出的.

拉格朗日法(Lagrange method)　利用拉格朗日函数(参见"拉格朗日鞍点"),把约束非线性规划问题转化为无约束极小化问题求解的一种方法.对于非线性规划问题

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t. } g_i(\boldsymbol{x})\leqslant 0\quad (i=1,2,\cdots,p),\end{cases}$$

设 $f(\boldsymbol{x}),g_i(\boldsymbol{x})(i=1,2,\cdots,p)$ 均具有一阶连续偏导数,将拉格朗日函数表示为上述非线性规划问题的目标函数与加权约束函数之和:

$$L(\boldsymbol{x},\boldsymbol{\lambda})=f(\boldsymbol{x})-\sum_{i=1}^{p}\lambda_i g_i(\boldsymbol{x}),$$

其中 $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_p)^{\mathrm{T}}$ 是拉格朗日乘子.令其各偏导数为零,得到一方程组,解此方程组可得到非线性规划问题的最优解.

非线性规划对偶函数(dual function of nonlinear programming)　非线性规划问题中一种由目标函数和约束函数构成的函数.对于非线性规划问题

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t. } g_i(\boldsymbol{x})\leqslant 0\quad (i=1,2,\cdots,p),\end{cases}$$

设 $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_p)^{\mathrm{T}}$,

$$L(\boldsymbol{x},\boldsymbol{\lambda})=f(\boldsymbol{x})+\sum_{i=1}^{p}\lambda_i g_i(\boldsymbol{x})$$

是它的拉格朗日函数,则称

$$G(\boldsymbol{\lambda})=\min_{x\in\mathrm{R}^n}L(\boldsymbol{x},\boldsymbol{\lambda})$$

是它的对偶函数或拉格朗日对偶函数.

非线性规划对偶问题(dual problem of nonlinear programming)　与非线性规划问题相对应的另一非线性规划问题.对于非线性规划问题

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t. } g_i(\boldsymbol{x})\leqslant 0\quad (i=1,2,\cdots,p),\end{cases}$$

其(拉格朗日)对偶问题为

$$\begin{cases}\max\limits_{\lambda\in D}\ G(\boldsymbol{\lambda})=\max\limits_{\lambda\in D}\min\limits_{x\in\mathrm{R}^n}\ L(\boldsymbol{x},\boldsymbol{\lambda}),\\ D=\{\boldsymbol{\lambda}\in\mathrm{R}^p\,|\,G(\boldsymbol{\lambda})\text{ 存在},\boldsymbol{\lambda}\geqslant\boldsymbol{0}\},\end{cases}$$

其中 $L(\boldsymbol{x},\boldsymbol{\lambda})$ 是原非线性规划问题的拉格朗日函数(参见"非线性规划的对偶函数").非线性规划的对偶问题也称为极小-极大对偶问题.

极小-极大对偶问题(min-max dual problem)即"非线性规划的对偶问题".

二次规划的解法(solution to quadratic programming)　利用库恩-塔克尔条件求解二次规划问题的方法.将库恩-塔克尔条件中(参见"库恩-塔克尔条件")的

$$\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{p}\lambda_i^*\nabla g_i(\boldsymbol{x}^*)+\sum_{j=1}^{q}\mu_j^*\nabla h_j(\boldsymbol{x}^*)=0$$

应用于二次规划,并用 $\boldsymbol{y}=(y_1,y_2,\cdots,y_{n+p})^{\mathrm{T}}$ 代替 $\boldsymbol{\lambda}^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_{n+p}^*)^{\mathrm{T}}$,得到

$$\sum_{k=1}^{n}c_{jk}x_k+\sum_{i=1}^{p}a_{ij}y_{n+i}+y_j=c_j\quad (j=1,2,\cdots,n);\tag{1}$$

引入松弛变量 x_{n+i},则二次规划问题中的约束条件为

$$\sum_{j=1}^{n}a_{ij}x_j-x_{n+i}+b_i=0\quad (b_i\geqslant 0,i=1,2,\cdots,p);\tag{2}$$

再将库恩-塔克尔条件中的 $\lambda_i^* g_i(\boldsymbol{x}^*)=0$ 应用于二次规划,并考虑到式(2),得到

$$x_j y_j=0\quad (j=1,2,\cdots,n+p);\tag{3}$$

还有

$$x_j\geqslant 0,\ y_j\geqslant 0\ (j=1,2,\cdots,n+p).\tag{4}$$

联立求解(1)和(2),若得到的解也满足式(3)和(4),则这样的解就是原二次规划问题的最优解.但在式(1)中,c_j 可能为正,也可能为负.为了便于求解,先引入人工变量 $z_j(z_j\geqslant 0)$,式(1)变为

$$\sum_{i=1}^{p}a_{ij}y_{n+i}+y_j+\sum_{k=1}^{n}c_{jk}x_k+\mathrm{sgn}(c_j)z_j=c_j\quad (j=1,2,\cdots,n),\tag{5}$$

其中 $\mathrm{sgn}(c_j)$ 为符号函数(当 $c_j\geqslant 0$ 时,$\mathrm{sgn}(c_j)=1$;当 $c_j<0$ 时,$\mathrm{sgn}(c_j)=-1$).得到初始基可行解:z_j

$=\mathrm{sgn}(c_j)$ $(j=1,2,\cdots,n)$, $x_{n+i}=b_i$ $(i=1,2,\cdots,p)$, $x_j=0$ $(j=1,2,\cdots,n)$, $y_j=0$ $(j=1,2,\cdots,n+p)$. 只有当 $z_j=0$ 时，才能得到原问题的解. 修正上述问题，得到

$$\begin{cases}\min\ \varphi(\boldsymbol{z})=\sum_{j=1}^{n}z_j,\\ \text{s.t.}\ \sum_{i=1}^{m}a_{ij}y_{n+i}+y_i-\sum_{k=1}^{n}c_{jk}x_k+\mathrm{sgn}(c_j)z_j\\ \qquad =c_j\ \ (j=1,2,\cdots,n),\\ \qquad \sum_{j=1}^{n}a_{ij}x_j-x_{n+i}+b_i=0\\ \qquad (i=1,2,\cdots,p),\\ \qquad x_j\geqslant 0\quad (j=1,2,\cdots,n+p),\\ \qquad y_j\geqslant 0\quad (j=1,2,\cdots,n+p),\\ \qquad z_j\geqslant 0\quad (j=1,2,\cdots,n).\end{cases}\tag{6}$$

此线性规划问题应满足式(3)，即对每一个 j，不能使 x_j 和 y_j 同时为基变量. 解线性规划问题(6)，若得到最优解 $(x_1^*,x_2^*,\cdots,x_{n+p}^*,y_1^*,y_2^*,\cdots,y_{n+p}^*,z_1=0,\cdots,z_n=0)^{\mathrm{T}}$，则 $(x_1^*,x_2^*,\cdots,x_n^*)^{\mathrm{T}}$ 就是原二次规划问题的最优解.

二次规划(quadratic programming)　一类典型的非线性规划(问题). 指目标函数是自变量的二次函数，而约束条件全是线性的非线性规划(问题). 二次规划问题是非线性规划中较规范的一类问题. 二次规划问题的数学模型为

$$\begin{cases}\min\ f(\boldsymbol{x})=\sum_{j=1}^{n}c_jx_j+\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}c_{jk}x_jx_k,\\ \text{s.t.}\ \sum_{j=1}^{n}a_{ij}x_j+b_i\geqslant 0\qquad (i=1,2,\cdots,p),\\ \qquad x_j\geqslant 0\qquad\qquad (j=1,2,\cdots,n),\end{cases}$$

其中

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}},$$
$$c_{jk}=c_{kj}(j=1,2,\cdots,n;k=1,2,\cdots,n),$$

目标函数 $f(\boldsymbol{x})$ 中的第二项为二次型. 若该二次型正定(或半正定)，则目标函数 $f(\boldsymbol{x})$ 为严格凸函数(或凸函数). 因二次规划的可行域为凸集，故这种规划又称为二次凸规划. 对于这类问题，库恩-塔克尔条件不仅是极小点的必要条件，而且也是充分条件.

凸规划对偶问题(dual problem of convex programming)　一类综合的对偶问题. 指线性规划、带凸性的非线性规划和二次规划的对偶问题的综合形式. 例如，若二次规划问题的矩阵形式是

$$\begin{cases}\min\ f(\boldsymbol{x})=\boldsymbol{cx}+\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}Q\boldsymbol{x},\\ \text{s.t.}\ A\boldsymbol{x}\leqslant \boldsymbol{b},\end{cases}$$

则其对偶问题是

$$\begin{cases}\max\ G(\boldsymbol{\lambda})=\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}Q\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda}\right),\\ \text{s.t.}\ A^{\mathrm{T}}\boldsymbol{\lambda}+\boldsymbol{c}^{\mathrm{T}}+Q\boldsymbol{x}=0,\\ \qquad \boldsymbol{\lambda}\geqslant \boldsymbol{0}.\end{cases}$$

几何规划(geometric programming)　一类特殊的非线性规划(问题). 利用著名的几何平均数不大于算术平均数，即

$$\sqrt[n]{x_1x_2\cdots x_n}\leqslant\frac{x_1+x_2+\cdots+x_n}{n},$$

这一不等式，通过几何规划可以把具有非线性约束的非线性规划问题，转化为具有线性约束的一个对偶问题. 非线性规划的每一个结果和计算方法几乎都可以用几何规划去阐明，许多最优化问题都能表达成几何规划或它的推广. 几何规划是由达芬(Duffin, R. J.)、彼得森(Peterson, E.)和泽内尔(Zener, C.)合著的《几何规划的理论及其应用》一书而得名.

几何规划的数学模型为

$$\begin{cases}\min\ g_0(\boldsymbol{x}),\\ \text{s.t.}\ g_k(\boldsymbol{x})\leqslant 1\quad (k=1,2,\cdots,l),\\ \qquad \boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}>0,\end{cases}$$

其中

$$g_k(\boldsymbol{x})=\sum_{i=m_k}^{n_k}c_i\prod_{j=1}^{n}x_j^{a_{ij}}\ \ (k=0,1,2,\cdots,l;c_i>0),$$

m_k 和 n_k 是用来表示相继顺序项的个数，$m_0=1,2,\cdots,n$ 表示目标函数的项数，$n_l=m$ 表示包括目标函数、约束函数所有项数的总项数，指数 a_{ij} 是任意实数.

分式规划(fractional programming)　目标函数是分式函数的非线性规划(问题). 其数学模型是

$$\begin{cases}\min\ f(\boldsymbol{x})=\dfrac{\boldsymbol{p}^{\mathrm{T}}\boldsymbol{x}+\alpha}{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{x}+\beta},\\ \text{s.t.}\ A\boldsymbol{x}=\boldsymbol{b},\\ \qquad \boldsymbol{x}\geqslant\boldsymbol{0},\end{cases}$$

其中 $\boldsymbol{p},\boldsymbol{q}\in\mathrm{R}^n$，$\alpha,\beta\in\mathrm{R}$，$A=(a_{ij})_{m\times n}$，$\boldsymbol{b}\in\mathrm{R}^m$，且对于任意满足约束条件的 $\boldsymbol{x}$，都有 $\boldsymbol{qx}+\beta\neq 0$.

分式规划问题的解法为：把分式规划化为线性规划，再求该线性规划问题的最优解，即求线性规划问题

$$\begin{cases}\min\ (\boldsymbol{p}^{\mathrm{T}}\boldsymbol{y}+\alpha\tau),\\ \text{s.t.}\ A\boldsymbol{y}-\tau\boldsymbol{b}=\boldsymbol{0},\\ \qquad \boldsymbol{q}^{\mathrm{T}}\boldsymbol{y}+\beta\tau=1,\\ \qquad \boldsymbol{y}\geqslant\boldsymbol{0},\ \tau\geqslant 0\end{cases}\tag{1}$$

的最优解，这里 $\boldsymbol{x}=\boldsymbol{y}/\tau(\tau\in\mathrm{R},\tau>0)$. 设其最优解为 $(\tilde{\boldsymbol{y}}^{\mathrm{T}},\tilde{\tau})^{\mathrm{T}}$，若有 $\tilde{\tau}>0$，则当原分式规划有使(分母)$\boldsymbol{qx}+\beta>0$ 的最优解时，$\tilde{\boldsymbol{x}}=\tilde{\boldsymbol{y}}/\tilde{\tau}$ 为所求的最优解；若分式规划的某可行解(或最优解)使 $\boldsymbol{qx}+\beta<0$，则原分式规划问题为

$$\begin{cases}\min \dfrac{-\boldsymbol{p}^{\mathrm{T}}\boldsymbol{x}-\alpha}{-\boldsymbol{q}^{\mathrm{T}}\boldsymbol{x}-\beta},\\ \text{s.t. } A\boldsymbol{x}=\boldsymbol{b},\\ \quad \boldsymbol{x}\geqslant \boldsymbol{0}.\end{cases}$$

其相应的线性规划问题为

$$\begin{cases}\min\ (-\boldsymbol{p}^{\mathrm{T}}\boldsymbol{y}-\alpha\tau),\\ \text{s.t. } A\boldsymbol{y}-\tau\boldsymbol{b}=\boldsymbol{0},\\ \quad -\boldsymbol{q}^{\mathrm{T}}\boldsymbol{y}-\beta\tau=1,\\ \quad \boldsymbol{y}\geqslant\boldsymbol{0},\tau\geqslant 0.\end{cases}\tag{2}$$

当未知 $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{x}+\beta$ 的符号时，须解线性规划问题(1)和(2).

多目标规划

多目标规划(multiple objective programming) 数学规划的一个较新的分支学科. 它研究多于一个的实值目标函数(向量目标函数)在给定约束条件下的极值(极大化或极小化)问题. 有限维极小化形式的多目标规划问题(VMP)表示为

$$V-\min_{x\in X}\boldsymbol{f}(\boldsymbol{x})=(f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x}))^{\mathrm{T}},$$

其中 $X\subset\mathbb{R}^n$ 是约束集，$\boldsymbol{f}: X\to\mathbb{R}^m$ 是(m 维)向量目标函数，$V-\min$ 表示“向量极小化”. 设(VMP)中的约束集由不等式约束和等式约束表示为 $X=\{\boldsymbol{x}\in\mathbb{R}^n \mid g_i(\boldsymbol{x})\leqslant 0, i=1,2,\cdots,p; h_j(\boldsymbol{x})=0, j=1,2,\cdots,q\}$，则问题(VMP)也记为

$$V-\begin{cases}\min\ (f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x}))^{\mathrm{T}},\\ \text{s.t. } g_i(\boldsymbol{x})\leqslant 0\quad (i=1,2,\cdots,p),\\ \quad h_j(\boldsymbol{x})=0\quad (j=1,2,\cdots,q).\end{cases}$$

多目标规划问题(VMP)也称多目标最优化问题，有时简称多目标规划. 若(VMP)中的 $f_k(\boldsymbol{x})$ $(k=1,2,\cdots,m)$，$g_i(\boldsymbol{x})$ $(i=1,2,\cdots,p)$ 和 $h_j(\boldsymbol{x})$ $(j=1,2,\cdots,q)$ 都是 $\boldsymbol{x}$ 的线性函数，则相应的问题称为多目标线性规划. 若(VMP)的 $f_k(\boldsymbol{x})$ $(k=1,2,\cdots,m)$，$g_i(\boldsymbol{x})$ $(i=1,2,\cdots,p)$ 和 $h_j(\boldsymbol{x})$ $(j=1,2,\cdots,q)$ 中有 $\boldsymbol{x}$ 的非线性函数，则称其为多目标非线性规划. 在多目标规划的研究中，变量空间和目标空间均可不限于有限维的情况. 对于这些空间是巴拿赫空间和拓扑向量空间的情形，也已建立了系统的数学理论. 多目标规划的主要起源是：经济学中埃奇沃思(Edgeworth, F. Y.)(1874 年)和帕雷托(Pareto, V.)(1906 年)关于均衡竞争和福利经济的研究. 康托尔(Cantor, G. (F. P.))(1895 年)和豪斯多夫(Hausdorff, F.)(1906 年)创立的有序空间数学理论，则为多目标规划的理论形成提供了重要的研究工具. 1951 年，库普曼斯(Koopmans, T. C.)结合数量经济的研究，给出多目标规划的一些基本概念和结果，引起人们的关注. 1958 年，赫维茨(Hurwicz, L.)开始把对多目标规划问题的研究扩展到一般的抽象空间之后，使这一学科逐渐发展成为数学规划的一门具有系统数学理论的新分支.

有效解(efficient solution) 亦称帕雷托有效解或帕雷托解. 多目标规划的基本概念之一. 对于多目标规划问题(VMP)(参见“多目标规划”)，设 $\tilde{\boldsymbol{x}}\in X$，若不存在 $\boldsymbol{x}\in X$ 使得 $f_k(\boldsymbol{x})\leqslant f_k(\tilde{\boldsymbol{x}})$ $(k=1,2,\cdots,m)$，并且其中至少有一为严格不等式，则称 $\tilde{\boldsymbol{x}}$ 是(VMP)的有效解. 多目标规划问题(VMP)的有效解必是它的弱有效解(参见“弱有效解”).

弱有效解(weakly efficient solution) 亦称帕雷托弱有效解. 多目标规划的基本概念之一. 对于多目标规划问题(VMP)(参见“多目标规划”)，设 $\tilde{\boldsymbol{x}}\in X$，若不存在 $\boldsymbol{x}\in X$ 使得 $f_k(\boldsymbol{x})<f_k(\tilde{\boldsymbol{x}})$ $(k=1,2,\cdots,m)$，则称 $\tilde{\boldsymbol{x}}$ 是(VMP)的弱有效解.

绝对最优解(absolute optimal solution) 亦称帕雷托最优解. 多目标规划的基本概念之一. 它是使多目标规划问题中的每一目标函数都达到最优值的解. 对于多目标规划问题(VMP)(参见“多目标规划”)，设 $\boldsymbol{x}^*\in X$，若对所有的 $k=1,2,\cdots,m$ 和任意的 $\boldsymbol{x}\in X$ 都有 $f_k(\boldsymbol{x}^*)\leqslant f_k(\boldsymbol{x})$，则称 $\boldsymbol{x}^*$ 是(VMP)的绝对最优解. 由于多目标规划问题的绝对最优解要求问题中的每一目标函数都在该处同时达到最优，这显然只有当问题是很特殊的情况时才会发生，因此对于一个给定的多目标规划问题，它的绝对最优解一般是不存在的.

帕雷托解(Pareto solution) 即“有效解”. 亦称帕雷托有效解.

非劣解(noninferior solution) 多目标规划的基本概念之一. 对于包括有定量和定性属性的多指标决策问题(参见“多目标决策问题”)，其非劣解是指在所给的可供选择的方案集中，已找不到使每一指标都能改进的解. 在多目标规划中，它即指有效解(参见“有效解”)和较多最优解(参见“较多最优解”).

较多有效解(majorly efficient solution) 多目标规划的基本概念之一. 是使多目标规划问题中的较多个目标函数为非劣的解. 对于多目标规划问题(VMP)(参见“多目标规划”)，设 $\tilde{\boldsymbol{x}}\in X$，若不存在 $\boldsymbol{x}\in X$ 使得 $f_k(\boldsymbol{x})\leqslant f_k(\tilde{\boldsymbol{x}})$ $(k=1,2,\cdots,m)$ 中至少有

$$\left[\frac{m}{2}\right]$$

个成立，并且其中至少有一为严格不等式，则称 $\tilde{\boldsymbol{x}}$ 是(VMP)的较多有效解. 多目标规划问题(VMP)的较多有效解必是它的有效解(参见“有效解”)和较多最优解(参见“较多最优解”).

较多最优解(majorly optimal solution) 多目标规划的基本概念之一. 是使多目标规划问题中的

较多个目标函数达到最优的解. 对于多目标规划问题(VMP)(参见"多目标规划"),设 $x^* \in X$,若对任意的 $x \in X$,不等式组 $f_k(x^*) \leqslant f_k(x)$ $(k=1,2,\cdots,m)$ 中至少有

$$\left[\frac{m}{2}\right]$$

个成立,则称 x^* 是(VMP)的较多最优解. 多目标规划问题(VMP)的较多最优解必是它的有效解(参见"有效解"). 此外,(VMP)的较多有效解必是它的较多最优解.

满意解(satisfaction solution) 多目标规划的基本概念之一. 是在两个或两个以上的有效解(参见"有效解")或非劣解(参见"非劣解")中,根据决策者的偏爱,从中选出的解.

多目标线性规划(multiobjective linear programming) 一类多目标规划. 指向量目标函数和约束函数都是决策变量的线性函数的多目标规划(问题). 多目标线性规划问题(VLP)的数学模型是

$$V - \begin{cases} \min\ (c_1^{\mathrm{T}}x, c_2^{\mathrm{T}}x, \cdots, c_m^{\mathrm{T}}x)^{\mathrm{T}}, \\ \text{s.t.}\ \ Ax \leqslant b, \\ \qquad x \geqslant 0, \end{cases}$$

其中 $x \in \mathrm{R}^n$, $c_k \in \mathrm{R}^n (k=1,2,\cdots,m)$ 和 $b \in \mathrm{R}^p$ 是已知向量, $A \in \mathrm{R}^{p\times n}$是已知矩阵. 设 $x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$, $c_k=(c_{k1},c_{k2},\cdots,c_{kn})^{\mathrm{T}}$, $b=(b_1,b_2,\cdots,b_p)^{\mathrm{T}}$ 和 $A=(a_{ij})_{p\times n}$,则(VLP)也即

$$\begin{cases} \min\ c_{11}x_1 + c_{12}x_2 + \cdots + c_{1n}x_n, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \min\ c_{m1}x_1 + c_{m2}x_2 + \cdots + c_{mn}x_n, \\ \text{s.t.}\ \ a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leqslant b_1, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \qquad a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pn}x_n \leqslant b_p, \\ \qquad x_j \geqslant 0 \quad (j=1,2,\cdots,n). \end{cases}$$

有效解存在性定理(existence theorem of efficient solution) 多目标规划理论的基本定理. 该定理断言:对于多目标规划问题(VMP),若约束集 $X \subset \mathrm{R}^n$ 是非空紧集,向量目标函数 $f: X \to \mathrm{R}^m$ 是连续函数,则(VMP)的有效解存在.

接触定理(contact theorem) 多目标规划理论中刻画有效解和弱有效解本质特性的定理. 即:设多目标规划问题(VMP)的约束集 $X \subset \mathrm{R}^n$ 非空,则:

1. $\tilde{x}$ 是(VMP)的有效解当且仅当
$$f(X) \cap (f(\tilde{x}) - \mathrm{R}_+^m) = \{f(\tilde{x})\}.$$
2. $\tilde{x}$ 是(VMP)的弱有效解当且仅当
$$f(X) \cap (f(\tilde{x}) - \mathrm{int}\mathrm{R}_+^m) = \varnothing,$$

其中 $\mathrm{R}_+^m = \{x \in \mathrm{R}^n | x \geqslant 0\}$.

标量化基本定理(scalarization basic theorem) 研究多目标规划理论和方法的重要基础和工具. 设多目标规划问题(VMP)的约束集 X 非空, $u: \mathrm{R}^m \to \mathrm{R}$ 是单调函数. 考虑与(VMP)相关的标量极小化问题(P_u):

$$\min_{x\in X} u(f(x)).$$

有如下的标量化基本定理:

1. 若 $u(y)$ 关于 $y \in \mathrm{R}^m$ 是严格增函数, $\tilde{x}$ 是(P_u)的最优解,则 $\tilde{x}$ 是(VMP)的有效解.

2. 若 $u(y)$ 关于 $y \in \mathrm{R}^m$ 是增函数, $\tilde{x}$ 是(P_u)的最优解,则 $\tilde{x}$ 是(VMP)的弱有效解.

评价函数方法(evaluation function method) 一类求解多目标规划问题的方法. 是通过构造一个把多个目标转化为单个数值目标的评价函数,将多目标规划问题归结为求解相应评价函数的数值最优化问题的方法的统称. 一般地,采用不同形式的评价函数可求得多目标规划问题的不同意义下的解. 常用的评价函数方法有线性加权和法、极大-极小法和理想点法等.

线性加权和法(linear weighted sum method) 一种评价函数方法. 是按各目标的重要性赋予它相应的权系数,然后对其线性组合进行寻优的求解多目标规划问题的方法. 对于问题(VMP),即归结为求

$$\min_{x\in X} \sum_{k=1}^{m} w_k f_k(x)$$

的最优解,其中 $w_k \geqslant 0 (k=1,2,\cdots,m)$ 为权系数,且

$$\sum_{k=1}^{m} w_k = 1.$$

极大-极小法(maximal-minimal method) 一种求解多目标规划问题的评价函数方法. 思想是:先对各目标函数作极大值选择作为评价函数,然后再在可行域上进行极小化求解. 即将问题(VMP)化为求

$$\min_{x\in X}\{\max_{1\leqslant k\leqslant m} f_k(x)\}$$

的最优解,或转化为求等价问题

$$\begin{cases} \min\ \lambda, \\ \text{s.t.}\ \ x \in X, \\ \qquad f_k(x) \leqslant \lambda \quad (k=1,2,\cdots,m) \end{cases}$$

的最优解.

理想点法(ideal point method) 一种评价函数方法. 使各目标值尽可能逼近其理想(最优)值的求解多目标规划问题的一种评价函数方法. 对于问题(VMP),设

$$f_k^* = \min_{x\in X} f_k(x)\ (k=1,2,\cdots,m),$$

则称 $f^* = (f_1^*, f_2^*, \cdots, f_m^*)^{\mathrm{T}}$ 是(VMP)的理想点. 方法是:将求解多目标规划问题(VMP)化为求数值最优化问题

$$\min_{x\in \mathrm{R}} \| f(x) - f^* \|$$

的最优解，其中 $\|\cdot\|$ 是 R^m 中某意义下的模.

交互规划算法(interactive programming algorithm)　一类求解多目标规划问题的方法. 指以分析者的求解和决策者的抉择交互进行的求解多目标规划问题的一类方法. 这是一类人机对话式的迭代求解过程，每一轮按分析求解所得的信息，提供给决策者作偏爱选择，反复进行，直至得到决策者认可的满意解. 常用的交互规划算法有逐步法(STEM)、权衡比替代法和逐次线性加权和法等.

逐步法(step method)　亦称逐步宽容约束法. 简称 STEM. 求解多目标规划问题的一种交互规划算法. 该方法包括以下两个阶段：在分析求解阶段，分析者用理想点法对问题求解(参见"理想点法")，并将所得解的对应目标值(称参考目标值)和理想目标值提供给决策者；在决策选择阶段，决策者在比较参考目标值和理想目标值的基础上，对已满意的目标给出使其目标值做出让步的宽容量，以换取使不满意目标得到改善，据此组成新的约束集后再求解. 如此交替进行，逐步求得决策者对各目标均满意的解. 此法适用于各目标函数是决策变量的线性函数和约束集是紧集的情况. 方法由贝拿扬(Benayoun, R.)等于 1971 年提出.

权衡比替代法(trade-off replacement method)　一种交互规划算法. 是一种求解带非线性目标函数和线性约束条件的多目标规划问题的交互规划算法. 这是一个包括两个阶段的迭代算法. 在决策选择阶段，决策者对两两目标权衡其得失，给出当前迭代点处各两两目标的权衡比. 在分析求解阶段，分析者借助求解带线性约束的非线性规划问题的 F-W 法(参见"F-W 法"和"近似线性化法")，并对其中隐含评价函数的边际置换率作权衡比替代，把问题归为求解一线性规划问题，求得下一迭代点. 如此反复进行，从而得到决策者愈来愈满意的解. 此法于 1972 年由杰夫里翁(Geoffrion, A. M.)等提出.

平方加权和法(weighted quadratic sum method)　一种求解多目标规划问题的方法. 指用平方和加权形式使各目标尽可能逼近其想象最好值的求解多目标规划问题的一种方法. 对于问题(VMP)，先求出各个目标的想象最好值(下界)$\hat{f}_k(k=1,2,\cdots,m)$，即

$$\hat{f}_k \leqslant \min_{x\in X} f_k(\boldsymbol{x})\quad (k=1,2,\cdots,m).$$

设 $w_k>0(k=1,2,\cdots,m)$ 为事先给定的一组权系数，且

$$\sum_{k=1}^{m} w_k = 1,$$

再求

$$\min_{x\in X}\sum_{k=1}^{m} w_k(f_k(\boldsymbol{x})-\hat{f}_k)^2$$

的最优解.

虚拟目标法(fictitious objective method)　一种求解多目标规划问题的方法. 指使各目标尽可能逼近其虚拟目标值的求解多目标规划问题的一种方法. 对于问题(VMP)，方法是先给出各目标的虚拟目标值 $f_k^0(k=1,2,\cdots,m)$，即

$$\min_{x\in X} f_k(\boldsymbol{x}) \geqslant f_k^0 \neq 0\ (k=1,2,\cdots,m),$$

再求

$$\min_{x\in X}\left\{\sum_{k=1}^{m}\left[\frac{f_k(\boldsymbol{x})-f_k^0}{f_k^0}\right]^2\right\}^{\frac{1}{2}}$$

的最优解.

约束法(constraint method)　一种求解多目标规划问题的方法. 指在多个目标函数中，确定一个主要目标，对其他各目标都给定一个所期望的值，将其转化为约束条件，从而把多目标最优化模型化为单目标最优化模型求解的方法.

功效系数法(efficiency coefficient method)　亦称功效函数法. 一种求解同时具有极小化和极大化目标函数的多目标规划问题的方法. 设前 r 个目标要求越小越好，而后 $m-r$ 个目标要求越大越好. 针对这些目标函数值的"好坏"用一个功效系数 d_k $(0\leqslant d_k\leqslant 1, k=1,2,\cdots,m)$ 表示它(或称"打分")，当 $f_k(\boldsymbol{x})$ 达到最满意值时令 $d_k=1$(或 $d_k\approx 1$)；达到最不满意值时令 $d_k=0$. 然后，由各功效系数构造出问题的评价函数进行求解(参见"评价函数方法"). 如可采用求极大化几何平均问题：

$$\max_{x\in X}\sqrt[m]{d_1(f_1(\boldsymbol{x}))d_2(f_2(\boldsymbol{x}))\cdots d_m(f_m(\boldsymbol{x}))},$$

设得最优解 $\bar{x}$，则它即为所考虑多目标规划问题的解. 功效函数 $d_k=d_k(f_k(\boldsymbol{x}))(k=1,2,\cdots,m)$ 的具体形式是多种多样的，常用的有线性型和指数型两种.

乘除法(multiplication division method)　一种求解多目标规划问题的方法. 对于同时具有极小化和极大化目标函数的多目标规划问题，设前 r 个目标 $f_k(\boldsymbol{x})(k=1,2,\cdots,r)$ 要求极小化，后 $m-r$ 个目标 $f_k(\boldsymbol{x})(k=r+1,\cdots,m)$ 要求极大化，则问题可归为求"乘除"目标形式的数值极小化问题

$$\min_{x\in X}\left(\prod_{k=1}^{r} f_k(\boldsymbol{x})\Big/\prod_{k=r+1}^{m} f_k(\boldsymbol{x})\right)$$

的最优解，其中 $f_k(\boldsymbol{x})>0(k=1,2,\cdots,m)$.

分层序列法(stratified sequencing method)　亦称优先级法. 一种求解多目标规划问题的方法. 指把多目标规划问题中的各个目标，按轻重缓急分成不同层次，然后按先后顺序求解的方法.

优先级法(priority order method)　即"分层序列法".

宽容分层序列法(forbearing stratified sequenc-

ing method）　一种求解多目标规划问题的方法．指取一组事先给定的宽容值，作为各相应目标最优值的允许偏差，与分层序列法（参见“分层序列法”）类似，逐层求解的方法．

目标规划（goal programming）　多目标规划的分支方向之一．指研究由决策者对多个目标各提供希望达到的目标值信息，并以各目标函数与对应目标值间的偏差总和极小化为数学模型的分支．目标规划的数学模型也简称目标规划，它是 20 世纪 60 年代初，由查恩斯（Charnes，A.）和库伯（Cooper，W. W.）首先提出，到 20 世纪 70 年代受到人们的重视并得到广泛应用．设有多目标规划问题

$$\min_{x\in X} \boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x}))^{\mathrm{T}},$$

其中$X\subset \mathbb{R}^n$是约束集．首先由决策者对各目标给出对应的目标值 $\hat{\boldsymbol{f}} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_m)^{\mathrm{T}}$，同时引入各目标函数关于其对应目标值的正、负偏差：

$$d_i^+ = \frac{1}{2}(|f_i(\boldsymbol{x}) - \hat{f}_i| + (f_i(\boldsymbol{x}) - \hat{f}))$$

$$(i = 1,2,\cdots,m);$$

$$d_i^- = \frac{1}{2}(|f_i(\boldsymbol{x}) - \hat{f}_i| - (f_i(\boldsymbol{x}) - \hat{f}_i))$$

$$(i = 1,2,\cdots,m),$$

则有目标规划模型

$$\begin{cases} \min \sum_{i=1}^{m} (w_i^+ d_i^+ + w_i^- d_i^-), \\ \text{s.t. } f_i(\boldsymbol{x}) - d_i^+ + d_i^- = \hat{f}_i \\ \qquad (i = 1,2,\cdots,m), \\ \qquad \boldsymbol{x} \in X, \\ \qquad d_i^+ \geqslant 0,\ d_i^- \geqslant 0 \\ \qquad (i = 1,2,\cdots,m), \end{cases}$$

其中 w_i^+ 和 w_i^-（$i=1,2,\cdots,m$）分别是正偏差和负偏差的权系数．注意，上述目标规划模型是一具 $n+2m$ 个决策变量的（单目标）数学规划问题．若将 m 个目标分成 L 个等级 $f_i^s(x)$（$s=1,2,\cdots,L;i=1,2,\cdots,ls$），以 $p_1, p_2, \cdots, p_L$ 表示等级记号，约定 p_1 优先于 p_2，p_2 优先于 p_3，…，即有

$$p_1 \gg p_2 \gg \cdots \gg p_L.$$

于是，可建立一般（分层）目标规划的数学模型如下：

$$L-\begin{cases} \min \left[p_s \sum_{i=1}^{ls} (w_{si}^+ d_{si}^+ + w_{si}^- d_{si}^-)\right]_{s=1}^{L}, \\ \text{s.t. } f_i^s(\boldsymbol{x}) - d_{si}^+ + d_{si}^- = \hat{f}_i^s \\ \qquad (s = 1,2,\cdots,L; i = 1,2,\cdots,ls), \\ \qquad \boldsymbol{x} \in X, \\ \qquad d_{si}^+ \geqslant 0,\ d_{si}^- \geqslant 0 \\ \qquad (s = 1,2,\cdots,L; i = 1,2,\cdots,ls), \end{cases}$$

其中 w_{si}^+ 和 w_{si}^- 分别是正偏差和负偏差的权系数，$\hat{f}_i^s$ 是目标函数 $f_i^s(\boldsymbol{x})$ 的目标值．

线性目标规划（linear goal programming）　一种目标规划（问题）．指目标函数和约束函数均为决策变量的线性函数的目标规划（问题）（参见“目标规划”）．线性目标规划问题的数学模型为

$$L-\begin{cases} \min \left[p_s \sum_{i=1}^{ls} (w_{si}^+ d_{si}^+ + w_{si}^- d_{si}^-)\right]_{s=1}^{L}, \\ \text{s.t. } \boldsymbol{c}_i^{s\mathrm{T}} \boldsymbol{x} - d_{si}^+ + d_{si}^- = \hat{f}_i^s \\ \qquad (s = 1,2,\cdots,L;\ i = 1,2,\cdots,ls), \\ \qquad A\boldsymbol{x} \leqslant \boldsymbol{b}\ (\text{或} \geqslant \boldsymbol{b}\ \text{或} = \boldsymbol{b}), \\ \qquad \boldsymbol{x} \geqslant \boldsymbol{0},\ d_{si}^+ \geqslant 0,\ d_{si}^- \geqslant 0 \\ \qquad (s = 1,2,\cdots,L; i = 1,2,\cdots,ls), \end{cases}$$

它的每一层次都是一个线性规划问题．线性目标规划的基本原理断言：

1. 线性目标规划必存在满意解，且可用单纯形法（参见“单纯形法”）求解其对应的线性规划问题而得到．

2. 若线性目标规划的单纯形表中任一检验列皆为零向量，或虽非零向量，但自上至下第一个非零分量为正数，则该单纯形表对应的基可行解为线性目标规划的满意解．

线性目标规划的基本原理（basic principle of linear goal programming）　见“线性目标规划”．

目标单纯形法（goal simplex method）　求解线性目标规划问题的基本方法．求解过程包括以下四个步骤：

1. 建立初始单纯形表（参见“单纯形表”）．

2. 判定最优解．基可行解为线性目标规划的满意解，而基可行解为线性规划的基本最优解，故用单纯形法（参见“单纯形法”）求解其对应的线性规划问题得到的基本可行解也是线性目标规划的满意解．

3. 换基迭代．若检验数不符合基本原理 2（参见“线性目标规划”）的要求，则：

1）确定进基变量．从左至右逐个检查每一检验数列，第一个不符合基本原理 2 要求的检验数列所对应的变量为进基变量．

2）确定出基变量和进行换基迭代的方法和过程完全与求解线性规划的单纯形法一样．求得新表后再返回步骤 2．根据线性规划的单纯形法中避免可行基循环的法则可知，经有限次换基迭代后必可求得线性目标规划的满意解．

4. 对满意解及其对目标达到情况进行分析．把对所求得的满意解及其对应目标的实现情况通报决策者．若决策者对此满意，则计算终止；否则，根据决策者的要求，调整目标的优先等级或优先权系数或重新修改预定目标值，然后再对修改后的模型重新求解，直到决策者满意为止．

无限维多目标规划问题（infinite dimensional

multiobjective programming problem) 目标是无限维的多目标规划问题. 设 $\mathscr{X}$ 和 $\mathscr{Y}$ 是抽象空间(Banach 空间、拓扑向量空间或线性空间), $X \subset \mathscr{X}$ 是非空集合, $\boldsymbol{f}: \mathscr{X} \to \mathscr{Y}$ 是映射, 则极小化形式的无限维多目标规划问题(TMP)的数学模型是

$$T - \min_{x \in X} \boldsymbol{f}(\boldsymbol{x}),$$

其中"$T-\min$"表示无限维极小化的意思, $\mathscr{Y}$ 中的序由 $\mathscr{Y}$ 中的凸锥确定. 当 $\mathscr{X}=\mathrm{R}^n$, $\mathscr{Y}=\mathrm{R}^m$, 并且 R^m 中用自然序时, 问题(TMP)即成为有限维多目标规划问题(VMP)(参见"多目标规划"). 在锥有效解的意义下(参见"锥有效解"), 自 20 世纪 70 年代以来, 已建立了无限维多目标规划的系统理论.

锥有效解(cone efficient solution) 无限维多目标规划问题的解. 考虑无限维多目标规划问题(TMP)(参见"无限维多目标规划问题"), 设 $K \subset \mathscr{Y}$ 是尖闭凸锥, $\mathrm{int}\, K \neq \varnothing$. 若 $\tilde{x} \in X$, 并且不存在 $\boldsymbol{x} \in X$, 使得 $\boldsymbol{f}(\tilde{\boldsymbol{x}}) - \boldsymbol{f}(\boldsymbol{x}) \in K \setminus \{\boldsymbol{0}\}$ $(\mathrm{int}\, K)$, 则称 $\tilde{x}$ 是问题(TMP)的 K(弱)有效解. 由于 K 是一锥, 故这类解也统称为锥有效解.

动 态 规 划

动态规划(dynamic programming) 数学规划的分支学科. 它以多阶段最优化问题为研究对象. 动态规划所研究的问题(数学模型)也简称动态规划. 20 世纪 50 年代初, 美国数学家贝尔曼(Bellman, R.)等把多阶段决策问题表示成一系列子问题, 而每个子问题是易于寻优的, 他们提出了解决多阶段决策问题的"最优性原理", 从而对许多由线性规划和非线性规划不易处理的问题可以用动态规划去求解. 特别地, 对于解决离散型问题, 它已成为非常有用的工具. 动态规划已被广泛地应用于工程技术、管理决策、国民经济及军事系统中的各种实际问题. 动态规划模型分为离散确定型、离散随机型、连续确定型和连续随机型四类.

阶段(stage) 动态规划术语. 把一个动态规划问题的全过程按某种方式适当地分成若干个相互联系的部分, 称其中每一部分为一阶段.

状态(state) 动态规划术语. 指动态规划问题中某阶段某支路的始点, 同时也是前一阶段某支路的终点. 通常一个阶段包含若干个状态. 用 x_k 表示第 k 阶段的某一状态, $x_k^{(i)}$ 表示在第 k 阶段的第 i 个状态, $X_k=\{x_k^{(1)}, x_k^{(2)}, \cdots, x_k^{(i)}, \cdots, x_k^{(r)}\}$ 表示第 k 阶段的状态集合. 这里所指的状态, 必须具有三个特性:

1. 能够用来描述受控过程的演变特征.

2. 无后效性. 若某阶段状态给定, 则在这阶段以后, 过程的发展不受前面各阶段状态的影响. 即过程的过去历史只能通过当前的状态去影响其未来的发展, 当前的状态就是未来过程的初始状态.

3. 可知性. 规定的各阶段状态变量的值, 不管是直接或间接的, 都应是可知的.

状态变量(state variable) 描述过程状态的变量(参见"状态").

控制变量(control variable) 亦称决策变量. 描述决策的变量. 常用 $u_k(x_k)$ 表示第 k 阶段当状态处于 x_k 时的控制变量, $D_k(x_k)$ 表示第 k 阶段的允许决策集合, 则 $u_k(x_k) \in D_k(x_k)$.

策略(policy) 动态规划的基本概念之一. 指由每一阶段的决策 $u_i(x_i)$ $(i=1,2,\cdots,n)$ 组成的决策函数序列, 记为

$$P_{1,n}(x_1)=\{u_1(x_1), u_2(x_2), \cdots, u_n(x_n)\}.$$

由第 k 阶段开始到终点为止的决策函数序列称为子策略, 记为

$$P_{k,n}(x_k)=\{u_k(x_k), u_{k+1}(x_{k+1}), \cdots, u_n(x_n)\}.$$

由第 k 阶段开始到终点的过程, 称为原过程的后部子过程或称为 k 子过程.

子策略(sub-policy) 见"策略".

最优策略(optimal policy) 动态规划的基本概念之一. 是从允许策略集合中找出达到最优效果的策略. 在实际问题中, 可供选择的策略有一定的范围, 称此范围为允许策略集合. 初始状态为 x_k 的后部子过程所有子策略中最优的策略称为最优子策略.

最优子策略(optimal sub-policy) 见"最优策略".

多阶段决策问题(multistage decision problem) 一类决策问题. 指可以分成若干相互联系的阶段的决策问题. 每个阶段都需要做出决策, 由各阶段的决策构成一个决策序列, 即策略(参见"策略"). 在允许策略中选择一个最优策略, 使预定标准的目标函数值达到最优.

指标函数(indicator function) 用来衡量所实现过程优劣的一种数量指标. 指在多阶段决策问题中, 定义在全过程和所有后部子过程上的一个确定数量函数. 在不同的问题中, 指标的含义也不同. 表示为

$$\begin{aligned} v_{k,n} &= v_{k,n}(x_k, u_k, x_{k+1}, \cdots, x_{n+1}) \\ &= \psi(x_k, u_k, v_{k+1,n}(x_{k+1}, \cdots, x_{n+1})), \end{aligned}$$

式中 $\psi(x_k, u_k, v_{k+1,n})$ 对其变元 $v_{k+1,n}$ 来说要严格单调. 常见的指标函数是取各阶段指标和的形式, 即

$$v_{k,n} = \sum_{j=k}^{n} v_j(x_j, u_j),$$

其中 $v_j(x_j, u_j)$ 表示第 j 阶段的指标. 因此, $v_{k,n} = v_k(x_k, u_k) + v_{k+1,n}(x_{k+1}, \cdots, x_{n+1})$. 指标函数是初始

状态和策略的函数，用 $v_{k,n}(x_k, P_{k,n}(x_k))$ 表示，即 $v_{k,n}(x_k, P_{k,n}) = v_k(x_k, u_k) + v_{k+1,n}(x_{k+1}, P_{k+1,n})$，其子策略 $P_{k,n}(x_k) = \{u_k(x_k), P_{k+1,n}(x_{k+1})\}$，或用 $v_{0,n-1}(x_0, P_{0,n-1})$ 表示，即

$$\begin{aligned} v_{0,n-1}(x_0, P_{0,n-1}) &= \sum_{i=0}^{n-1} v_i(x_i, u_i) \\ &= \sum_{i=0}^{k-1} v_i(x_i, u_i) + \sum_{i=k}^{n-1} v_i(x_i, u_i) \\ &= v_{0,k-1}(x_0, P_{0,k-1}(x_0)) \\ &\quad + v_{k,n-1}(\tilde{x}_k, P_{k,n-1}(\tilde{x}_k)), \end{aligned}$$

其中 $v_i(x_i, u_i)$ 是第 i 阶段的指标函数，$v_{0,k-1}$ 和 $v_{k,n-1}$ 分别是上述两个子过程的总指标. $\tilde{x}_k = T_{k-1}(x_{k-1}, u_{k-1})$ 是第 $k-1$ 阶段到第 k 阶段的状态转移规律.

最优指标值(optimal indicator value)　动态规划的基本概念之一. 是指标函数的最优值. 用 $f_k(x_k)$ 表示，即

$$\begin{aligned} f_k(x_k) &= v_{k,n}(x_k, P^*_{k,n}(x_k)) \\ &= \operatorname*{opt}_{P_{k,n}} v_{k,n}(x_k, P_{k,n}(x_k)), \end{aligned}$$

其中"opt"是最优化的记号，根据问题的要求可取极小化(min)或极大化(max).

最短路线问题(shortest path problem)　一个多阶段决策问题. 即选择一条由始点到终点的路线，使总距离最短的问题. 它的性质是：若最短路线在中间站通过某点，则从该点出发到达终点的这条路线，对于从该点出发到达终点的所有可能选择的不同路线中是最短的. 寻找最短路线的方法是：从最后一段开始，由后向前逐步递推，找出各中间点到终点的最短路线，最后找出由始点到终点的最短路线.

最优性原理(principle of optimality)　亦称贝尔曼原理. 动态规划的基本原理. 该原理是：无论过去的状态和决策如何，对前面的决策所形成的状态而言，余下的诸决策必须构成最优策略. 此性质是对整个过程的最优化策略而言的. 这个原理是美国数学家贝尔曼(Bellman, R.)于 1957 年出版的《动态规划》一书中提出的.

贝尔曼原理(Bellman principle)　即"最优性原理".

动态规划方法(dynamic programming method)　一种求解动态规划问题的方法. 指从终点逐段向始点方向寻找最优策略的方法. 应用贝尔曼原理，用一个基本的递推关系式，从终点开始按倒过来的顺序逐段向始点方向寻找最优途径. 即把原问题分成许多比它简单而又相互联系的子问题，在每个子问题的求解中，都利用它的一个后部子问题的最优化结果，依次进行，最后一个子问题所得到的最优解就是原问题的最优解. 因此，动态规划方法是既把当前一段和未来各段分开，又把当前效益和未来效益结合起来考虑的一种最优化方法.

动态规划的函数基本方程(functional basic equation of dynamic programming)　用递推关系表示动态规划的基本方程. 即

$$f_k(x_k) = \min_{u_k(x_k)} \{d_k(x_k, u_k(x_k)) + f_{k+1}(u_k(x_k))\}$$
$$(k = n, n-1, \cdots, 2, 1),$$
$$f_{n+1}(x_{n+1}) = 0$$

或写成

$$f_k(x_k) = d_k(x_k, G),$$

这里 $f_k(x_k)$ 表示由某段状态 x_k 到终点 G 的最短距离，k 表示由状态 x_k 到终点 G 还有 k 段要走.

动态规划的分析函数方程(analytic functional equation of dynamic programming)　用分析函数表达的动态规划的基本方程. 即

$$f_k(x) = \min_{u_k(x)} \{d_k(x, u_k(x)) + f_{k+1}(u_k(x))\},$$

其中 x 是状态变量，决策 $u_k(x)$ 是第 k 阶段状态 x 的函数，$d_k(x, u_k(x))$ 表示在第 k 阶段由点 x 到点 $u_k(x)$ 的距离，$f_k(x)$ 表示由第 k 阶段点 x 到终点的最短距离.

标号法(labeling method)　寻找最短路线问题的最优方案的一个有效方法. 基本步骤为：

1. 先标出离终点最近的一段，将距离数字分别写在该点上方的方格内.

2. 在标下一段时，正要标号的某点到该段已标号的各点的距离，分别加上已标号点的数字而取其中最小者，这就是某点到终点的最短距离. 将距离数字填入某点上方方格内，并用直线连结起来表示某点到终点的最短路线.

3. 继续按逆推过程计算，直至始点为止.

这里是从后向前标的，也可以从前向后标.

逆序解法(inverse order method)　一种求解动态规划问题的方法. 指以起点为始端，终点为终端的右行解法. 以终点为始端，起点为终端的左行解法称为顺序解法. 逆序解法和顺序解法只表示行进方向的不同或始端与终端的颠倒，但都是在行进方向规定后，从最后一段向前逆推计算，逐段找出最优途径.

顺序解法(order method)　见"逆序解法".

状态转移方程(state transition equation)　动态规划的基本方程. 是表示由 k 阶段到 $k+1$ 阶段的状态转移规律的方程，用 $x_{k+1} = T_k(x_k, u_k)$ 表示之.

动态规划的基本方程(basic equation of dynamic programming)　逆序解法的基本方程和顺序解法的基本方程的总称. 逆序解法的基本方程是

$$\begin{aligned} f_k(x_k) &= \operatorname*{opt}_{P_{k,n}} v_{k,n}(x_k, P_{k,n}) \\ &= \operatorname*{opt}_{u_k} v_{k,n}(x_k, u_k, P^*_{k+1,n}) \end{aligned}$$

$$= \underset{u_k \in D_k(x_k)}{\text{opt}} \{v_k(x_k, u_k) + f_{k+1}(x_{k+1})\}$$

$$(k = n, n-1, \cdots, 2, 1),$$

$$f_{n+1}(x_{n+1}) = 0,$$

其中 opt 表示最优化. 其递推过程为从 $k=n$ 开始，逐段向前推移，直至求出 $f_1(x_1)$ 为止. 顺序解法的基本方程是

$$f_k(x_k) = \underset{u_{k-1} \in D_{k-1}^r(x_k)}{\text{opt}} \{v_{k-1}(x_k, u_{k-1}) + f_{k-1}(x_{k-1})\} \ (k=2,3,\cdots,n+1),$$

$$f_1(x_1) = 0,$$

其中 $x_{k-1} = T_{k-1}^r(x_k, u_{k-1})$. 状态转移方程为 $x_k = T_k^r(x_{k+1}, u_k)$. 递推过程为从 $k=2$ 开始，逐段向后推移，直至求出 $f_{n+1}(x_{n+1})$ 为止.

允许策略(permissible policy) 动态规划中一种两不同阶段间可能存在的策略.

$$P_{0,n-1} = \{u_0(x_0), u_1(x_1), \cdots, u_{n-1}(x_{n-1})\}$$

表示从 0 到 $n-1$ 阶段的一个允许策略，$P_{0,n-1}(x_0)$ 表示由 x_0 出发，从 0 阶段到 $n-1$ 阶段所有允许策略的集合. 允许策略为最优策略的充分必要条件是：对任意的 k $(0<k<n-1)$ 和 $x_0 \in X_0$(第 0 阶段的状态集)，有

$$v_{0,n-1}(x_0, P_{0,n-1}^*) = \underset{P_{0,k-1} \in P_{0,k-1}(x_0)}{\text{opt}} \{v_{0,k-1}(x_0, P_{0,k-1}) + \underset{P_{k,n-1} \in P_{k,n-1}(\tilde{x}_k)}{\text{opt}} v_{k,n-1}(\tilde{x}_k, P_{k,n-1})\},$$

其中 opt 表示最优化. 当 v 是效益函数时，opt 取 max；当 v 是损失函数时，opt 取 min. 若允许策略 $P_{0,n-1}^*$ 是最优策略，则对任意的 $k(0<k<n-1)$，它的子策略 $P_{k,n-1}^*$ 对于以 $x_k^* = T_{k-1}(x_{k-1}^*, u_{k-1}^*)$ 为起点的 k 到 $n-1$ 子过程来说，必是最优策略.

函数迭代法(function iteration method) 亦称函数空间迭代. 动态规划的求解方法之一. 是以段数作为参变数，先求在各个不同段数下的最优策略，然后从对应的最优解中选出最优者，从而同时确定了最优段数. 具体做法是：先选定一初始函数 $f_1(i)$，$f_1(i) = c_{iN}(i=1,2,\cdots,N-1)$，$f_1(N)=0$；然后用下列递推关系求出 $\{f_k(i)\}$：

$$f_k(i) = \min_j [c_{ij} + f_{k-1}(j)]$$

$$(i = 1, 2, \cdots, N-1),$$

$$f_k(N) = 0 \quad (k > 1).$$

这里 $f_k(i)$ 表示由 i 点出发，不重复地朝固定点走 k 步后的最短路线(不一定到达点 N). 若在 k 步到达点 N，则可停止，这时 $f_k(i) = f_{k+1}(i) = \cdots = f(i)$. 若在 k 步未达到点 N，则 $f_k(i)$ 不是由点 i 至点 N 的最优函数，而是由 i 至 $f_k(i)$ 所到达的点的最优函数，此时，

$$\lim_{k \to \infty} f_k(i) = f(i).$$

策略迭代法(policy iteration method) 动态规划的求解方法之一. 是先给出初始策略 $\{u_0(i)\}(i=1,2,\cdots,N-1)$，然后按某种方式求得新策略 $u_1(i)$，$u_2(i)$，…，直至最终求出最优策略的方法. 具体做法是：

1. 先选一无回路的初始策略 $\{u_0(i)\}(i=1,2,\cdots,N-1)$，$u_0(i)$ 表示在此策略下由点 i 到达的下一个点.

2. 由策略 $u_k(i)$ 求指标函数 $f_k(i)$，即由方程组

$$f_k(i) = c_{i,u_k(i)} + f_k(u_k(i))$$

$$(i = 1, 2, \cdots, N-1),$$

$$f_k(N) = 0 \quad (k = 0, 1, \cdots)$$

解出 $f_k(i)$，其中 $c_{i,u_k(i)}$ 为已知.

3. 由指标函数 $f_k(i)$ 求策略 $u_{k+1}(i)$，$u_{k+1}(i)$ 是 $\min\{c_{i,N} + f_k(u_k)\}$ 的解.

4. 按步骤 2 和 3 反复迭代，可逐次求得 $\{u_k(i)\}$ 和 $\{f_k(i)\}$，一直找到某一 k，使 $u_k(i) = u_{k-1}(i)$ 对所有 i 成立，则 $u_k(i)$ 就是最优策略，其相应的 $f_k(i)$ 为最优值，并且 $\{f_k(i)\}$ 一致收敛于方程组

$$f(i) = \min_j \{c_{ij} + f(j)\}$$

$$(i = 1, 2, \cdots, N-1),$$

$$f(N) = 0 \quad (c_{N,N} = 0)$$

的解. 在实际应用中，策略迭代法比函数迭代法要好些.

资源分配问题(resources allocation problem) 一类可用动态规划方法求解的非线性规划问题. 即把总量为 S 的某种资源，用于生产 n 种产品的问题. 若把数量为 y_j 的资源分配于生产第 j 种产品，其收益为 $g_j(y_j)$，问如何分配，才使生产 n 种产品的总收益最大? 其数学模型为

$$\begin{cases} \max Z = \sum_{j=1}^{n} g_j(y_j), \\ \text{s.t.} \ \sum_{j=1}^{n} y_j = S \quad (S \text{ 为正整数}), \\ y_j \geqslant 0 \qquad (j = 1, 2, \cdots, n). \end{cases}$$

一维离散化方法(one dimensional discrete method) 解资源分配问题的动态规划方法. 当 y 在 $[0,S]$ 上连续变动时，对区间 $[0,S]$ 进行分割，令 $y = 0, \Delta, 2\Delta, \cdots, m\Delta = S$，$f_k(y)$ 只在分割点上取值，把递推关系式

$$f_k(y) = \max_{0 \leqslant y_k \leqslant y} \{g_k(y_k) + f_{k+1}(y - y_k)\}$$

$$(k = n-1, \cdots, 1),$$

$$f_n(y) = g_n(y)$$

化为

$$\max_{p=1,2,\cdots,q} [g_k(p\Delta) + f_{k+1}(y - p\Delta)]$$

$$(k = n-1, \cdots, 1),$$

$$f_n(y) = g_n(y),$$

其中 $y=q\Delta$. 依次算出 $f_n(y), f_{n-1}(y), \cdots, f_1(y)$ 在 $y=0,\Delta,2\Delta,\cdots,m\Delta=S$ 上的值，再从 $f_1(S), f_2(S-\overline{y}_1), \cdots, f_k(a-\overline{y}_1-\overline{y}_2-\cdots-\overline{y}_{k+1})$ 求出 $\overline{y}_1,\overline{y}_2,\cdots,\overline{y}_k,\cdots$，直至最后求出 $\overline{y}_n$，则 $\overline{y}_1,\overline{y}_2,\cdots,\overline{y}_n$ 即为所求的最优分配值，$f_1(S)$ 为最大总收益.

二维资源分配问题(two dimensional resources allocation problem)　一类可用动态规划方法求解的非线性规划问题. 即把总量各为 a 和 b 的某种资源用于生产 n 种产品的问题. 若把数量为 x_i 和 y_i 的资源分配于生产第 i 种产品，其收益为 $g_i(x_i,y_i)$，问如何分配，才使生产 n 种产品的总收益最大?其数学模型为

$$\begin{cases}\max\ [g_1(x_1,y_1)+g_2(x_2,y_2)+\cdots+g_n(x_n,y_n)],\\ \text{s.t.}\ \ x_1+x_2+\cdots+x_n=a,\\ \qquad y_1+y_2+\cdots+y_n=b,\\ \qquad x_i\geqslant 0,\ y_i\geqslant 0\ \ (i=1,2,\cdots,n).\end{cases}$$

用动态规划方法来解，则设状态变量为 (x,y)，其中 x 和 y 分别表示分配用于生产第 k 种产品至第 n 种产品的第一种和第二种原料的数量. 设决策变量为 (x_k,y_k)，x_k 和 y_k 分别表示分配给第 k 种产品用的第一种和第二种原料的数量. 状态转换关系为：$\tilde{x}=x-x_k$，$\tilde{y}=y-y_k$，式中 $\tilde{x}$ 和 $\tilde{y}$ 分别表示用来生产第 $k+1$ 种产品至第 n 种产品的第一种和第二种原料的数量. 允许决策集合为

$$D_k(x,y)=\left\{u_k\ \middle|\ \begin{matrix}0\leqslant x_k\leqslant x,\\ 0\leqslant y_k\leqslant y\end{matrix}\right\}.$$

$f_k(x,y)$ 表示以第一种原料数量为 x，第二种原料数量为 y，分配用于生产第 k 种产品至第 n 种产品时所得到的最大收入. 由最优性原理(参见“最优性原理”)得

$$\begin{aligned}f_n(x,y)&=g_n(x,y),\\ f_k(x,y)&=\max_{\substack{0\leqslant x_k\leqslant x,\\ 0\leqslant y_k\leqslant y}}[g_k(x_k,y_k)\\ &\quad+f_{k+1}(x-x_k,y-y_k)]\\ &\quad(k=n-1,\cdots,1).\end{aligned}$$

此时，可采用离散化的方法来计算.

二维离散化方法(two dimensional discrete method)　解二维资源分配问题的动态规划方法. 即把矩形 $0\leqslant x\leqslant a, 0\leqslant y\leqslant b$ 分成网格，在格子点上进行计算. 若把 x 和 y 各分为 m_1 和 m_2 等分，则总共有 $(m_1+1)(m_2+1)$ 个格点，对每个 k 需要计算的 $f_k(x,y)$ 共有 $(m_1+1)(m_2+1)$ 个. 采取以时间换空间的方法，如拉格朗日乘数法、逐次逼近法和疏密法等. 它们都是以增加计算的时间来实现减少内存量的方法.

拉格朗日乘子法(Lagrange multiplier method)　一种求二维资源分配问题的降维方法. 它把二维资源分配问题降为一维分配问题. 具体做法是：把二维分配问题化为

$$\begin{cases}\max\ \{g_1(x_1,y_1)+g_2(x_2,y_2)+\cdots\\ \qquad +g_n(x_n,y_n)-\lambda(y_1+y_2+\cdots+y_n)\},\\ \text{s.t.}\ \ x_1+x_2+\cdots+x_n=a,\\ \qquad x_i\geqslant 0,\ y_i\geqslant 0\ \ (i=1,2,\cdots,n),\end{cases}$$

其中 λ 为拉格朗日乘子，是一个固定的参数. 令

$$h_i(x_i)=h_i(x_i,\lambda)=\max_{y_i\geqslant 0}[g_i(x_i,y_i)-\lambda y_i],$$

为了使此式有意义，可设

$$\lim_{y_i\to\infty}\frac{g_i(x_i,y_i)}{y_i}=0,$$

于是问题变为

$$\begin{cases}\max\ [h_1(x_1)+h_2(x_2)+\cdots+h_n(x_n)],\\ \text{s.t.}\ \ x_1+x_2+\cdots+x_n=a,\\ \qquad x_i\geqslant 0\ \ (i=1,2,\cdots,n).\end{cases}$$

这是一个一维分配问题，可用对一维的方法去求解. 这里，由于 λ 是参数，因此最优解 $\overline{x}_i$ 是参数 λ 的函数，相应的 $\overline{y}_i$ 也是 λ 的函数，即 $x_i=\overline{x}_i(\lambda)$，$y_i=\overline{y}_i(\lambda)$ 为其解. 若

$$\sum_{i=1}^{n}\overline{y}_i(\lambda)=b,$$

则 $\{\overline{x}_i,\overline{y}_i\}$ 为原问题的最优解；如果

$$\sum_{i=1}^{n}\overline{y}_i(\lambda)\neq b,$$

利用插值法调整 a 的值，直到满足

$$\sum_{i=1}^{n}\overline{y}_i(\lambda)=b$$

为止. 对于高维的问题也可用此法来降低维数.

双约束资源分配问题(double restriction resource allocation problem)　一种带两类约束条件的最优化问题. 指同时考虑物资分配限制和投资能力限制的最优化问题. 考虑物资总量为 N，投资能力为 M，生产 n 种产品的问题. 设把数量为 y_j 的物资分配于生产第 j 种产品，其收益为 $g_j(y_j)$，投资额为 $k_j(y_j)$，则其数学模型是

$$\begin{cases}\max\ \displaystyle\sum_{j=1}^{n}g_j(y_j),\\ \text{s.t.}\ \ \displaystyle\sum_{j=1}^{n}y_j\leqslant N\qquad (\text{物资约束}),\\ \qquad \displaystyle\sum_{j=1}^{n}k_j(y_j)\leqslant M\ \ (\text{投资约束}),\\ \qquad y_j=1,2,\cdots,N.\end{cases}$$

逐次逼近法(successive approximation method)　一种求二维资源分配问题的降维方法. 即把二维资源分配问题逐次降为一维分配问题. 具体做法是：选取初始点 $\boldsymbol{x}^0=\{x_1^0,x_2^0,\cdots,x_n^0\}$，满足

$$\sum_{i=1}^{n} x_i^0 = a.$$

固定 $\boldsymbol{x}$ 为 $\boldsymbol{x}^0$，先对 $\boldsymbol{y}$ 求解，把二维分配问题变为一维问题，即用一维方法求解问题

$$\begin{cases} \max \ \{g_1(x_1^0, y_1) + g_2(x_2^0, y_2) + \cdots \\ \qquad + g_n(x_n^0, y_n)\}, \\ \text{s.t.} \ \ y_1 + y_2 + \cdots + y_n = b, \\ \qquad y_i \geqslant 0 \quad (i = 1, 2, \cdots, m). \end{cases}$$

设其最优解为 $\boldsymbol{y}^0 = \{y_1^0, y_2^0, \cdots, y_n^0\}$，再固定 $\boldsymbol{y}$ 为 $\boldsymbol{y}^0$，对 $\boldsymbol{x}$ 求解，即求解问题

$$\begin{cases} \max \ \sum_{i=1}^{n} g_i(x_i, y_i^0), \\ \text{s.t.} \ \ \sum_{i=1}^{n} x_i = a, \\ \qquad x_i \geqslant 0 \quad (i = 1, 2, \cdots, n). \end{cases}$$

设其最优解为 $\boldsymbol{x}^1 = \{x_1^1, x_2^1, \cdots, x_n^1\}$，再固定 $\boldsymbol{x}$ 为 $\boldsymbol{x}^1$，对 $\boldsymbol{y}$ 求解. 这样逐次进行，得到一系列的解 $\{\boldsymbol{x}^k\}$，$\{\boldsymbol{y}^k\}(k=0,1,2,\cdots)$，且

$$\sum_{i=1}^{n} g_i(x_i^0, y_i) \leqslant \sum_{i=1}^{n} g_i(x_i^0, y_i^0) \leqslant \sum_{i=1}^{n} g_i(x_i^1, y_i^0).$$

序列

$$\left\{ \sum_{i=1}^{n} g_i(x_i^k, y_i^k) \right\}$$

是单调上升的，一般只收敛到某一局部最优解.

疏密法(sparse dense method)　求解动态规划问题的方法. 是逐步缩小区域加密网格的方法. 先用少数的格子点，在求出某一局部最优解后，再在最优解附近的小范围内进一步细分，并求出在细分格子点上的某一局部最优解. 如此继续细分下去，直到满足要求为止. 应用此法时要结合对指标函数特性的了解进行分析，以防止漏网.

嵌套分配问题(nesting allocation problem)　一类特殊的资源分配问题. 是存在两个或两个以上层次的分配关系的一类较复杂的资源分配问题(参见“资源分配问题”). 其数学模型为

$$\begin{cases} \max \ \sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij}(a_{ij}), \\ \text{s.t.} \ \ \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \leqslant M \quad \begin{pmatrix} i = 1, 2, \cdots, m, \\ j = 1, 2, \cdots, n \end{pmatrix} \ \text{(总量约束)}, \\ \qquad \sum_{i=1}^{m} a_{ij} \leqslant N_j \qquad (j = 1, 2, \cdots, n) \quad \text{(分量约束)}, \end{cases}$$

其中整数 $a_{ij} \geqslant 0$ 表示分配给第 j 个部门用于第 i 个方面的物资数，$Z_{ij}(a_{ij})$ 表示相应的利润，M 表示供分配的总物资数，N_j 表示分配给第 j 个部门物资量的限额.

生产与存贮问题(production and inventory problem)　一个多阶段决策问题. 其最优目标是正确制定生产计划、确定各个时期的产量，使在几个时期内的生产成本费用和存贮费用之和最小. 其约束条件是在满足市场对该产品的需要量的同时，使库存量在整个计划期末为零.

复合系统工作可靠性问题(reliability problem of compound system)　一个多阶段决策问题. 它研究整个系统的成本、重量、体积等在某些条件的限制下，应如何选择各部件的备用元件数，使整个系统的工作可靠性最大?它有静态和动态两种模型. 静态模型是

$$\begin{cases} \max \ \prod_{i=1}^{N} p_i(z_i), \\ \text{s.t.} \ \ \sum_{i=1}^{N} c_i z_i \leqslant c, \\ \qquad \sum_{i=1}^{N} w_i z_i \leqslant w, \\ \qquad z_i \geqslant 0 \quad (i = 1, 2, \cdots, N), \end{cases}$$

其中 $i=1,2,\cdots,N$ 为部件，z_i 为 i 部件的备用元件数，c_i 为 i 部件中每个元件的费用，w_i 为 i 部件的重量，$p_i(z_i) \leqslant 1$ 为处于正常工作状态的概率，是 z_i 的严格单调上升函数. 动态模型是

$$f_k(x_k, y_k) = \max_{u_k \in D_k(x_k, y_k)} \{p_k(u_k) f_{k+1}(x_k - u_k c_k, y_k - u_k w_k)\},$$

$$f_{N+1}(x_{N+1}, y_{N+1}) = 1,$$

其中允许决策集合为

$$D_k(x_k, y_k) = \left\{u_k \,\middle|\, 0 \leqslant u_k \leqslant \min\left\{\frac{x_k}{c_k}, \frac{y_k}{w_k}\right\},\ u_k \text{ 为非负整数}\right\},$$

指标函数为

$$V_{k,N} = \prod_{i=k}^{N} p_i(u_k) \quad (1 \leqslant k \leqslant N),$$

状态转移方程为

$$x_{k+1} = x_k - u_k c_k,$$
$$y_{k+1} = y_k - u_k w_k \quad (1 \leqslant k \leqslant N),$$

状态变量 x_k 表示由第 k 个到第 N 个部件所容许使用的总费用，状态变量 y_k 表示由第 k 个到第 N 个部件所容许具有的总重量. 决策变量 u_k 为部件 k 上装的备用元件数 z_k(即 $u_k = z_k$).

排序问题(sequencing problem)　一个多阶段决策问题. 指在加工工件所经过的工序、各种工件在各道工序加工需要的时间为给定的条件下，如何确定加工工件的顺序，使得总的加工时间最短的问题. 最优排序规则是：

1. 找出工件 $i(1 \leqslant i \leqslant n)$ 在两台机床上加工时间 a_i, b_i 中的最小数.

2. 若最小数为 $a_k(1\leqslant k\leqslant n)$，则将 k 排在第一位，并从工件集合中去掉这个工件.

3. 若最小数为 b_k，则将工件 k 排在最后一位，并从工件集合中去掉这个工件.

4. 对剩下的工件重复上述手续，直至工件集合为空集时停止.

概括起来，其基本思路是：尽量减少机床等待加工的时间.

设备更新问题(equipment replacement problem)　一个多阶段决策问题. 它是要确定在计算期内应进行多少次更新，每次更新应在哪一年进行，以使在所考虑的计算期内的总利润最大(或总费用最小)的问题. 设 $f_k(S_k)$ 表示第 k 年至第 n 年内，期初有一台役龄为 S_k 的设备采用最优更新策略所能获得的最大利润额，$t_{n+1}(S_{n+1})$ 表示第 n 年底(第 $n+1$ 年初)役龄为 S_{n+1} 的设备残值，则基本递推方程是

$$f_{n+1}(S_{n+1}) = t_{n+1}(S_{n+1}),$$

$$f_k(S_k) = \max_{d_k\in\{K,R\}}\{v_k(S_k,d_k) + f_{k+1}(S_{k+1})\}$$

$$(k = n, n-1, \cdots, 1),$$

其中决策变量 d_k 表示第 k 年初对役龄为 S_k 的机器采用的决策，它只能取两个值，更新(用 R 表示)或继续使用(用 K 表示)，即 $D_k(S_k)=\{R,K\}$. 状态转移方程为

$$S_{k+1} = \begin{cases} S_{k+1} & (d_k = K), \\ 1 & (d_k = R). \end{cases}$$

设 $r_k(S_k)$ 表示第 k 年初役龄为 S_k 的设备继续使用一年的年收入，$u_k(S_k)$ 表示第 k 年初役龄为 S_k 的设备继续使用一年的年维修费(包括因维修而减少生产所引起的损失)，$c_k(S_k)$ 表示第 k 年初对役龄为 S_k 的设备进行更新时的一次性以旧换新的费用(购买和安装新设备的费用与旧机器残值之差)，$v_k(S_k, d_k)$ 表示第 k 年的年利润，则有

$$v_k(S_k,d_k) = \begin{cases} r_k(S_k) - u_k(S_k) & (d_k = K), \\ r_k(0) - u_k(0) - c_k(S_k) & (d_k = R). \end{cases}$$

关于设备更新问题，除上述情况外，还有损耗规律为确定性时的更新方案和失效规律为随机性时的更新方案，对于后者只能利用概率的方法进行研究. 下面着重介绍前者. 例如，机器损耗或者衰老的维持费不断增长，何时更新才使现值总费用达到最小?其折现值后的总费用为

$$K_n = \frac{A + \sum\limits_{i=1}^{n} c_i\left(\frac{1}{1+r}\right)^{i-1}}{1 - \left(\frac{1}{1+r}\right)^{n}},$$

式中 A 表示买一台新机器的价格，c_i 表示第 i 年的维持费，r 表示年利率，

$$\frac{1}{\left(1+\frac{1}{r}\right)^{i-1}}$$

表示现值系数，

$$c_i\left(\frac{1}{1+r}\right)^{i-1}$$

表示现值，

$$A + \sum_{k=1}^{i} c_k\left(\frac{1}{1+r}\right)^{k-1}$$

表示设备原值与现值累计之和. 列出维持费、现值系数、现值、设备原值＋现值累计和折现值后总费用 K_i 的表，进行比较，即可求出最佳更新年份 m 应是使现值总费用 K_m 达到最小值. 经化简后即要求

$$c_{m+1} = \frac{\left[A+\sum\limits_{i=1}^{m} c_i\left(\frac{1}{1+r}\right)^{i-1}\right]\left(1-\frac{1}{1+r}\right)}{1-\left(\frac{1}{1+r}\right)^{m}},$$

或者第一个使等式右端小于左端的 m. 又如，设已知时刻 t 时维持费用函数为 $x(t)$，求最佳更新时刻 T. T 期间总费用为

$$A + \int_0^T x(t)\mathrm{d}t,$$

单位时间内平均总费用为

$$y(T) = \frac{1}{T}\left\{A + \int_0^T x(t)\mathrm{d}t\right\},$$

则

$$A = Ty(T) - \int_0^T x(t)\mathrm{d}t.$$

因某些设备使 A 不断贬值，故 A 亦为时间的函数 $A(T)$，则

$$y(T) = \frac{1}{T}\left\{A(T) + \int_0^T x(t)\mathrm{d}t\right\},$$

$$A(T) - T\frac{\mathrm{d}A(T)}{\mathrm{d}T} = Tx(T) - \int_0^T x(t)\mathrm{d}t.$$

满足后式的时刻 T 是最佳处理掉此设备的时间.

随 机 规 划

随机规划(stochastic programming)　数学规划的分支学科. 它研究含有随机变量的数学规划问题(参见“数学规划”). 随机规划所研究的问题有时也简称随机规划. 由于随机现象广泛存在，随机规划问题在许多领域出现，特别是在管理科学、环境科学、水利、农业等方面有广泛的应用. 用不同的原则处理数学规划问题中的随机变量，导致不同类型的随机规划问题，主要的类型有：分布问题，多阶段有补偿问题和概率约束规划问题. 由于随机变量的介入，这些问题都比较复杂而不易求解. 因此，这一学科的发展既依赖于、也推动了一般数学规划理论的发展. 随机规划起始于 20 世纪 50 年代. 最初，丹齐

克(Dantzig,G. B.)和查恩斯(Charnes,A.)建立了几种随机规划模型. 20 世纪 60 年代末,贝雷模(Bereamu,B.)研究了分布问题解的公式,韦茨(Wets,R.)刻画了二阶段问题目标函数的性质,普里考帕(Prekopa,A.)对概率约束规划可行解的凸性作了研究,他们的工作为随机规划奠定了理论基论. 求解随机规划问题的方法有卡尔(Kall. P.)的逼近方法和厄莫里叶夫(Ermoliev,Y.)的随机拟次梯度法等.

决策原则(decision principle) 处理随机规划问题中随机变量的原则. 这里的决策是指形成某类形式的规划问题和给出最优解. 主要有两种决策原则:一种是等待观察到随机变量的实现后再作决策(文献中称为 wait and see);另一种是在观察到随机变量的实现之前作决策(文献中称为 here and now). 两种不同类型的决策原则导致不同类型的随机规划问题.

分布问题(distribution problem) 随机规划问题中目标函数最优值(或最优解)的概率分布问题. 在观察到随机规划问题中随机变量的实现以后,就得出一个确定性的数学规划问题. 随机变量不同的实现值对应于不同的确定性数学规划问题和它们相应的最优值(或最优解). 求这些最优值(或最优解)的概率分布的问题即称为分布问题. 设 $\boldsymbol{x}\in\mathrm{R}^n$ 是实向量,$\boldsymbol{c}(\omega)$,$A(\omega)$,$\boldsymbol{b}(\omega)$和 $\zeta(\omega)$依次是 n 维随机向量、$n\times m$ 随机矩阵、m 维随机向量和随机向量,$f(\cdot,\cdot)$和 $\boldsymbol{g}(\cdot,\cdot)$分别是实值函数和向量函数,对于线性随机规划问题(LSP):

$$\begin{cases} z(\omega)=\min\ \boldsymbol{c}^{\mathrm{T}}(\omega)\boldsymbol{x}, \\ \text{s.t.}\quad A(\omega)\boldsymbol{x}=\boldsymbol{b}(\omega), \\ \qquad\quad \boldsymbol{x}>\boldsymbol{0}, \end{cases}$$

和非线性随机规划问题(NSP):

$$\begin{cases} z(\omega)=\min\ f(\boldsymbol{x},\zeta(\omega)), \\ \text{s.t.}\quad \boldsymbol{g}(\boldsymbol{x},\zeta(\omega))\leqslant\boldsymbol{0}, \end{cases}$$

其分布问题即为求其(随机)最优值 $z(\omega)$(或最优解 $\tilde{\boldsymbol{x}}(\omega)$)的概率分布和其他与分布有关量的问题.

多阶段有补偿问题(multistage problem with recourse) 一类随机规划问题. 设数学规划问题的约束条件中有随机变量,且决策变量要在观察到随机变量的实现之前做出,则不论决策变量取什么值,一般都不会使约束条件对任何样本值都满足. 为此,必须引进补偿使约束条件能够满足. 求由补偿引起的最小损失的问题称为第 k 阶段问题,由此产生的问题称为多阶段有补偿问题(或多阶段有补偿规划).

1. 二阶段有补偿问题. 对于(LSP)(参见“分布问题”)中的约束条件,设补偿量为 $\boldsymbol{y}$,使

$$A(\omega)\boldsymbol{x}+W(\omega)\boldsymbol{y}=\boldsymbol{b}(\omega)\quad(\boldsymbol{y}\geqslant\boldsymbol{0}),$$

其中 $W(\omega)$是随机变量. 选取 $\boldsymbol{y}$,使引进补偿所产生的损失达到最小,即求解

$$\begin{cases} Q(\boldsymbol{x},\omega)=\min\ \boldsymbol{q}^{\mathrm{T}}\boldsymbol{y}, \\ \text{s.t.}\quad W(\omega)\boldsymbol{y}=\boldsymbol{b}(\omega)-A(\omega)\boldsymbol{x}, \\ \qquad\quad \boldsymbol{y}\geqslant\boldsymbol{0}. \end{cases}\tag{1}$$

考虑到原来的损失函数与由补偿引起的损失,最后得到如下问题:

$$\begin{cases} \min\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}+EQ(\boldsymbol{x},\omega), \\ \text{s.t.}\quad \boldsymbol{x}\geqslant\boldsymbol{0}, \end{cases}\tag{2}$$

其中 $\boldsymbol{c}\in\mathrm{R}^n$,$E$ 表示数学期望运算. 这一问题称为二阶段补偿问题,其中(1)称为第二阶段问题;(2)实质上是一个非线性规划问题,但其目标函数与随机变量有关. 1955 年,丹齐克(Dantzig,G. B.)在将线性规划用于飞机航班设计问题时发现客流量是随机的,于是首先建立了二阶段有补偿问题.

2. 多阶段有补偿问题. 若第二阶段问题中的 $\boldsymbol{y}$ 还需满足其他一些约束条件,那些约束条件中也含有随机变量,且 $\boldsymbol{y}$ 亦需在那些随机变量的观察值得到以前做出决策,这将导致三阶段有补偿问题. 依此类推,则可产生多阶段有补偿问题. 多阶段有补偿问题的一般形式为

$$\begin{cases} \min\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}+E[\min\ \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{y}_1]+E[\min\ \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{y}_2]+\cdots \\ \qquad +E[\min\ \boldsymbol{q}_t^{\mathrm{T}}\boldsymbol{y}_t]\cdots], \\ \text{s.t.}\quad W_i(\omega)\boldsymbol{y}_i=\boldsymbol{b}_i-A_i(\omega)\boldsymbol{y}_{i-1} \\ \qquad\quad (\boldsymbol{y}_i\geqslant\boldsymbol{0},i=1,2,\cdots,t), \\ \qquad\quad W(\omega)\boldsymbol{y}_1=\boldsymbol{b}_1-A_1(\omega)\boldsymbol{x}\quad(\boldsymbol{y}_1\geqslant\boldsymbol{0}), \\ \qquad\quad \boldsymbol{x}\in X, \end{cases}$$

其中 X 是与随机变量无关的集合. 极值问题

$$\begin{cases} \min\ \boldsymbol{q}_k^{\mathrm{T}}\boldsymbol{y}_k, \\ \text{s.t.}\quad W_k(\omega)\boldsymbol{y}_k=\boldsymbol{b}_k-A_k(\omega)\boldsymbol{y}_{k-1}\quad(\boldsymbol{y}_k\geqslant 0) \end{cases}$$

称为第 k 阶段问题.

概率约束规划(probabilistic constrained programming) 约束条件中含有随机变量的随机规划问题. 如果数学规划问题的约束条件中含有随机变量(参见“数学规划”),而决策要在观察到随机变量的实现之前做出,但只要求约束条件得到满足的概率不小于某一指定数,这种问题称为概率约束规划. 它的一般形式为

$$\begin{cases} \min\ f(\boldsymbol{x}), \\ \text{s.t.}\quad P(g_i(\boldsymbol{x},\zeta(\omega))\geqslant 0)\geqslant\alpha_i \\ \qquad\quad (i=1,2,\cdots,m), \\ \qquad\quad \boldsymbol{x}\in X \end{cases}\tag{1}$$

或

$$\begin{cases} \min\ f(\boldsymbol{x}), \\ \text{s.t.}\quad P(g_i(\boldsymbol{x},\zeta(\omega))\geqslant 0, \\ \qquad\quad i=1,2,\cdots,m)\geqslant\alpha, \\ \qquad\quad \boldsymbol{x}\in X, \end{cases}\tag{2}$$

其中 X 为与随机变量无关的集合.(1)中的概率约束称为单个概率约束,(2)中的概率约束称为联合概率约束.(1)和(2)也是非线性规划问题.1958 年,查恩斯(Charnes,A.)等人在研究石油加工厂库存容量设计问题时,建议在满足外界需求(随机)的概率不低于某一水平下进行最优设计,导致概率约束规划的研究.

最优值的可测性(measurability of optimum) 随机规划的基本概念之一.指分布问题(LSP)和(NSP)(参见“分布问题”)中最优值 $z(\omega)$ 的可测性.因为 $z(\omega)$ 随着样本点 ω 而变化,欲求其概率分布,首先要确认 $z(\omega)$ 是从概率空间 (Ω,F,P) 到博雷尔可测空间 (R,B) 的可测函数.1985 年,韦茨(Wets,R.)和洛克菲勒(Rockafellar,R.)证明了:若分布问题中的目标函数和约束条件中的函数均为下半连续的可测函数,则其最优值 $z(\omega)$ 也是可测函数.对于其最优解 $\bar{x}(\omega)$(一般是多值函数),可选出一个可测选择(measurable selection).在较严格条件下的可测性结果由卡尔(Kall,P.)于 1971 年和 1974 年得出.

决定区域(decision region) 随机规划的基本概念之一.对某一样本点 ω,若随机线性规划问题(LSP)的最优值有限,则一定可以用一最优可行基 B_i 表示为

$$z(\omega)=\boldsymbol{c}^{\mathrm{T}}B_i(\omega)B_i^{-1}(\omega)\boldsymbol{b}(\omega).$$

所有可以用此 B_i 表示出最优值的样本点的集合 $\widetilde{D}_i$ 称为对应于基 B_i 的决定区域.找出各个最优可行基对应的决定区域,即可得到 $z(\omega)$ 的分布函数

$$\begin{aligned}F_z(t)&=P(z(\omega)<t)\\&=\sum_{i=1}^{K}P(z(\omega)<t,\omega\in D_i)\\&=\sum_{i=1}^{K}P(\boldsymbol{c}^{\mathrm{T}}B_i(\omega)B_i^{-1}(\omega)\boldsymbol{b}(\omega)<t,\omega\in D_i),\end{aligned}$$

其中

$$D_i=\widetilde{D}_i\setminus(\bigcup_{j=1}^{i-1}\widetilde{D}_j).$$

完备补偿矩阵(complete recourse matrix) 多阶段有补偿问题中的矩阵.设第二阶段问题(1)(参见“多阶段有补偿问题”)中的补偿矩阵 W 为常数矩阵,且方程组 $E\boldsymbol{y}=\boldsymbol{e}$ 对 $\boldsymbol{e}$ 的任何值都有非负解 $\boldsymbol{y}$ 存在,则称 W 为完备补偿矩阵.具有完备补偿矩阵的第二阶段问题(1)的可行解集总非空,由补偿引起的损失函数 $Q(\boldsymbol{x},\omega)$ 总有限.这对保证二阶段有补偿问题目标函数的有限性起决定性作用.矩阵 $W=(I,-I)$(I 表示单位矩阵)是特别简单的完备补偿矩阵,称为简单补偿矩阵.具有简单补偿矩阵的二阶段有补偿问题的目标函数比较容易求出.

简单补偿矩阵(simple recourse matrix) 见“完备补偿矩阵”.

基分解法(basic decomposition method) 求解具有离散随机变量的二阶段有补偿问题的方法.原是一种求解具有特殊结构的大规模线性规划问题的方法,它把基于维数极大的可行基的迭代运算分解为对维数很小的矩阵的迭代运算.用它来求解具有离散随机变量的二阶段有补偿问题时,问题(2)(参见“多阶段有补偿问题”)的形式为

$$\begin{cases}\min\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}+\sum\limits_{j=1}^{N}P_j\,\boldsymbol{q}^{\mathrm{T}}\boldsymbol{y}_j,\\ \text{s.t. } A_j\boldsymbol{x}+W\boldsymbol{y}_j=\boldsymbol{b}_j\quad(j=1,2,\cdots,N),\\ \quad\boldsymbol{x}\geqslant\boldsymbol{0},\\ \quad\boldsymbol{y}_j\geqslant\boldsymbol{0}.\end{cases}$$

它的对偶问题的约束系数矩阵为

$$B=\begin{bmatrix}W & & -W & & I & \\ & 0 & & 0 & & 0\\ & 0 & & 0 & & 0\\ & W & & -W & & I\\ A_1 & A_K & -A_1 & -A_k & & I\end{bmatrix}.$$

从 B 中选取一可行基,它在每一块行 $(W,-W,I)$ 中的截取记为 $B_i(i=1,2,\cdots,N)$,把线性规划的迭代运算(计算检验数,换基)分解到各个 B_i 上进行.此法有很好的计算效率,较好地解决了具有离散随机变量的二阶段有补偿问题.此法由斯特拉茨基(Strazicky,B.)提出,并由卡尔(Kall,P.)作了改进.

对数拟凹测度(logarithmic quasi-concave measure) 一类特殊测度.设 P 是 n 维博雷尔可测空间 (R^n,B^n) 上的测度.若对任意的 $\lambda\in(0,1)$ 有

$$P(\lambda A+(1-\lambda)B)\geqslant[P(A)]^{\lambda}[P(B)]^{1-\lambda}\quad(\forall A,B\subset B^n),$$

其中 $\lambda A+(1-\lambda)B=\{\boldsymbol{t}\mid\boldsymbol{t}=\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y},\boldsymbol{x}\in A,\boldsymbol{y}\in B\}$,则称 P 是 (R^n,B^n) 上的对数拟凹测度.关于对数拟凹测度与概率约束规划可行解的凸性,有下列结论:

1. 若一测度 P 为对数拟凹测度,则它所对应的分布函数的对数为凹函数.

2. 若一随机变量的分布密度函数 $f(\boldsymbol{t})$ 可写为

$$f(\boldsymbol{t})=\gamma\mathrm{e}^{-Q(t)}$$

的形式(其中 γ 为正常数,$Q(\boldsymbol{t})$ 为凸函数),则它在 (R^n,B^n) 上诱导的概率测度为对数拟凹测度.

3. 若 ξ 为具有对数拟凹分布的随机向量,$g_i(i=1,2,\cdots,m)$ 为凹函数,则

$$G(\boldsymbol{x})=P(g_i(\boldsymbol{x},\xi)\geqslant 0,i=1,2,\cdots,m)$$

为对数凹函数,从而集合 $\{\boldsymbol{x}\mid G(\boldsymbol{x})\geqslant\alpha\}$ 为凸集.

4. 常用的正态分布、β 分布、维夏特分布为对数拟凹分布.

对数拟凹分布的理论及其在概率约束规划中的应用的研究由普里考帕(Prekopa,A.)完成.

拟凹测度(quasi-concave measure)　一类特殊测度. 设 P 是 n 维博雷尔可测空间(R^n, B^n)上的测度. 若对任意的 $\lambda\in(0,1)$有

$$P(\lambda A+(1-\lambda)B)\geqslant\min\{P(A),P(B)\}\quad(\forall A,B\in B^n),$$

则称 P 是(R^n,B^n)上的拟凹测度. 关于拟凹测度有以下结论:

1. 拟凹测度对应的分布函数为拟凹函数.

2. 若随机向量分布密度函数 $f(t)$使$(f(t))^{-1/m}$为凸函数,则

$$P(c)=\int_c f(t)\,\mathrm{d}t$$

定义的测度 P 为拟凹测度.

3. 若 ξ 具有拟凹概率分布,g_i 为凹函数,则

$$G(\boldsymbol{x})=P\{g_i(\boldsymbol{x},\xi)\geqslant 0,\ i=1,2,\cdots,m\}$$

为拟凹函数,从而相应的概率约束规划的可行解集$\{\boldsymbol{x}\mid G(\boldsymbol{x})\geqslant\alpha\}$为凸集.

4. t 分布、F 分布为拟凹概率分布.

拟凹概率分布及其在概率约束规划中的应用的研究由博雷尔(Borel,C.)于 1975 年完成.

随机拟次梯度法(stochastic quasi-subgradient method)　一个求解随机规划问题的算法. 满足下列条件的随机向量 ξ^k 称为凸函数$f(\boldsymbol{x})$在点 $\boldsymbol{x}^k$ 处的随机拟次梯度:

$$E(\xi^k/\boldsymbol{x}^0,\boldsymbol{x}^1,\cdots,\boldsymbol{x}^k)=\partial f(\boldsymbol{x}^k)+\boldsymbol{b}^k\quad(k=0,1,2,\cdots),$$

其中 $\partial f(\boldsymbol{x}^k)$为 f 在 $\boldsymbol{x}^k$ 处的次梯度的一个元素,$\{\boldsymbol{b}^k\}$为收敛于零的随机向量序列. 用迭代公式

$$\boldsymbol{x}^{k+1}=\pi_S(\boldsymbol{x}^k-\rho_k\gamma_k\xi^k(\omega))\tag{1}$$

求 $f(\boldsymbol{x})$在集合 S 上的极小值的算法称为随机拟次梯度法,其中 $\pi_S(\boldsymbol{y})$为 $\boldsymbol{y}$ 到 S 上的投影运算,ρ_k 为步长,γ_k 为规范化常数,$\xi^k(\omega)$为 ξ^k 的一个样本值.

厄莫里叶夫(Ermoliev,Y.)提出并详尽研究了随机拟次梯度法. 该方法的收敛性定理为:设用迭代公式(1)得出的点列$\{\boldsymbol{x}^k\}$满足条件

$$\|\boldsymbol{x}^k\|\leqslant L,\ E\{\|\xi^k\|^2/\boldsymbol{x}^0,\boldsymbol{x}^1,\cdots,\boldsymbol{x}^k\}\leqslant\eta_k^2\leqslant C_L\quad(k=1,2,\cdots),$$

其中 η_k 为随机变量,L 为常数. 若步长 ρ_k 满足条件

$$\sum_{k=1}^{\infty}\rho_k=\infty,\quad\sum_{k=1}^{\infty}\rho_k^2<\infty,$$

则$\{\boldsymbol{x}^k\}$几乎处处收敛于 $\operatorname{Arg}\min\limits_{x\in S}f(\boldsymbol{x})$中的点.

用随机拟次梯度法解二阶段有补偿问题(2)(参见"多阶段有补偿问题")时,选用 $c+u(\boldsymbol{x}^k,\omega_k)A(\omega_k)$作为上述公式(1)中 $\xi^k(\omega)$的样本值,这里 $u(\boldsymbol{x}^k,\omega_k)$为第二阶段补偿问题

$$\begin{cases}\min\ \boldsymbol{q}^{\mathrm{T}}\boldsymbol{y},\\ \text{s.t. } W(\omega_k)\boldsymbol{y}=\boldsymbol{b}(\omega_k)-A(\omega_k)\boldsymbol{x}_k,\\ \qquad \boldsymbol{y}\geqslant\boldsymbol{0}\end{cases}$$

的拉格朗日乘子. 可见,用随机拟次梯度法求解二阶段有补偿问题很方便. 步长 ρ_k 的选取则凭经验进行,ρ_k 选得好,点列$\{\boldsymbol{x}^k\}$可收敛得快些. 用随机拟次梯度法求出的迭代点列$\{\boldsymbol{x}^k\}$不能保证其对应函数值序列$\{f(\boldsymbol{x}^k)\}$的单调下降性. 由于这一点,该方法所得的结果有时不很理想.

求解随机规划的逼近方法(approximation method for solving stochastic programming)　求解随机规划问题的基本方法. 指用较为简单的随机变量代替随机规划问题中的随机变量,从而得到其近似解的方法. 一般皆用取有限个离散值的随机变量来代替连续分布的随机变量,从而使求期望和求概率的积分运算变成求和运算.

1. 二阶段有补偿问题的逼近方法. 用取有限个离散值的随机变量 ξ^k 代替原随机变量 ξ,相应的二阶段有补偿问题变为

$$\begin{cases}\min\ f_k(\boldsymbol{x})=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}+EQ(\boldsymbol{x},\xi^k),\\ \text{s.t. } \boldsymbol{x}\in X.\end{cases}$$

关于其最优解 $\tilde{\boldsymbol{x}}^k$ 的收敛性有如下结果:设 W 为完备补偿矩阵,ξ 为平方可积随机变量. 当$\{\xi^k\}$均方收敛于 ξ 时,$\{\tilde{\boldsymbol{x}}^k\}$的任一聚点均为原问题的最优解,若确有$\{\tilde{\boldsymbol{x}}^k\}$的一个收敛子序列$\{\tilde{\boldsymbol{x}}^{k_j}\}$存在,则其最优值序列收敛于原问题的最优值. 上述逼近问题等价于一个线性规划问题,故 $\tilde{\boldsymbol{x}}^k$ 容易求出. 当 k 充分大时,$\tilde{\boldsymbol{x}}^k$ 可作为原问题的近似解.

2. 概率约束规划的逼近方法. 用取离散值的随机变量 ξ^k 代替原约束条件中的随机变量,得到逼近问题

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t. } \boldsymbol{x}\in S_k=\{\boldsymbol{x}\mid P(A\boldsymbol{x}\geqslant\xi^k)\geqslant\alpha\},\\ \qquad \boldsymbol{x}\in X\end{cases}$$

的最优解 $\tilde{\boldsymbol{x}}^k$. 关于最优解的收敛性有如下结果:设 $f(\boldsymbol{x})$为连续函数,$\{\xi^k\}$分布收敛于 ξ,则$\{\tilde{\boldsymbol{x}}^k\}$的任一聚点必是原问题

$$\begin{cases}\min\ f(\boldsymbol{x}),\\ \text{s.t. } \boldsymbol{x}\in S=\{\boldsymbol{x}\mid P(A\boldsymbol{x}\geqslant\xi)\geqslant\alpha\},\\ \qquad \boldsymbol{x}\in X\end{cases}$$

的最优解. 上述逼近问题的可行解集为几个凸多面体之和集,因此该逼近问题也容易求解.

优　选　法

优选法(optimum seeking method)　亦称试验最优化方法. 一类实用的直接最优化方法. 指对于生产和科研中的选优问题(参见"选优问题"),利用数学原理,合理地安排试验点,使之较快地找到最优点或近似最优点的一类方法. 研究优选法的学科称为

优选学.优选法中比较成熟和著名的方法是解决单因素选优问题的斐波那契法(参见"斐波那契法")和近似黄金分割法(参见"0.618法"和"近似黄金分割法").在目标为单峰函数的条件下,这两种方法在减少试验次数方面是最优的方法.对于这一点,1953年美国的基弗(Kiefer,J.C.)给出了证明,1956年约翰逊(Johnson,H.T.)从动态规划的思路给出了另一证明,以后欧美和中国一些学者又进一步给出另一些证明.在斐波那契法的基础上又给出每批同时做几个试验的方法,如"时延"的试验方法、布尔察诺法等.华尔特(Wilde,D.J.)把优选法分为两类:一类是消去法,即通过试验将最优值所在的范围逐步缩小,直到缩小到要求的精度为止;另一类是爬山法,如爬山一样,从已知的信息逐步向更佳的方向移动,使目标函数值逐步上升(对极大化问题),直到不能再改进目标值为止.单因素优选法的核心是比较与鉴别.0.618法和斐波那契分数法的比较对象是两个试验点上的试验结果;对分法的比较对象是试验点上的试验结果与标准;分批试验法的比较对象是每批试验中的所有试验结果;爬山法的比较对象是前后两个试验点上的试验结果.1970年起,中国数学家华罗庚曾在中国推广了优选法.

试验最优化方法(experiment optimization method) 即"优选法".

选优问题(select excellency problem) 选择有关因素的最优(佳)点的问题.在生产实践和科学实验中,人们为了达到优质、高产、低消耗等目的,需要对有关因素的最优(佳)点进行选择.称所有这些选择最优(佳)点的问题为选优问题.

单因素优选法(univariate optimum seeking method) 寻求只有一个因素的选优问题最优点的方法.指在安排试验时,只考虑一个对目标影响最大的因素(其他因素看做固定不变),进行合理安排,找到最优点或近似最优点,以期达到最好的试验结果的方法.其数学描述是:应用此法,迅速找到一元目标函数的最大(或最小)值及其相应的最大(或最小)点.若目标函数是多峰的,则采取以下措施:

1.开始找一个峰,若达到要求,先采用,然后再找其他高峰.

2.先做一批分布得比较均匀的试验,若发现有多个峰,则在每个可能出现高峰的范围内做试验,把这些峰都找出来.

第一批分布均匀的试验点可用满足 $\alpha:\beta=0.618:0.382$的 α 和 β.这样,有峰的区间即(α,β)或(β,α),而已做过的试验结果下一次仍都可利用.有时把峰确定在较小范围以后,也常用抛物线插值法(参见"抛物线插值法"),取抛物线的顶点近似峰值点.

对分法(bisection method) 一种单因素优选法.是把试验区间对分为两半的一种试验方法.首先,根据经验确定试验区间(a,b).第一次试验在(a,b)的中点 $x_1=(a+b)/2$ 处做.若试验结果表明 x_1 处取值过大(或过小)了,则去掉大于 x_1(或 x_1 以下)的一半,第二次试验在(a,x_1)的中点 $x_2=(a+x_1)/2$(或在(x_1,b)的中点 $x_2=(x_1+b)/2$)处做.做了第一次试验,可将区间(a,b)缩小一半,然后在保留区间中点做第二次试验,再根据第二次试验结果,又将区间缩小一半.如此继续在余下的区间内试验下去,直至找到所要求的点为止.只有符合下面两个条件,才能应用对分法:

1.要有一个标准(或具体指标),否则无法鉴别试验结果的好坏.

2.要预知该因素对指标的影响规律,即能够从一个试验的结果直接分析出该因素的值是取大了还是取小了.否则,无法取舍哪段区间,即无从下手做下一次试验.

在有些问题中,只须鉴别标准,这时可以在该因素的高点和低点分别作两个试验,若能显示为不同的质量差异,可以以此为判别原则而应用对分法.

平分法(halver method) 亦称二分法.即"对分法".

波尔察诺法(Bolzano method) 即"对分法".原来的方法是用于求在已知区间内目标函数为单调函数的根的.它选优的速度比斐波那契法(参见"斐波那契法")更快.

对半法(half and half method) 一种单因素优选法.方法是每次在留下试验区间的中点的邻域同时安排两个尽量靠近的试验,对这两个试验进行比较,然后决定去掉右半区间还是左半区间,如此继续试验下去.

分批试验法(block search) 一种单因素优选法.此方法是每一批多做几个试验,同时进行比较,这样可以减少检验的时间或代价,一批一批地试验下去,直至找到最优点为止.若试验结果需要较长的时间才能得到,或者检验一次需要花很大代价,而且每次同时检验多个样品几乎和检验一个样品所花的时间或代价相近,则采用分批试验法为好.分批试验法不同于斐波那契法(参见"斐波那契法").分批试验法分为均分分批试验法和比例分割分批试验法两种.

均分分批试验法(share out equally block search) 一种分批试验法.指每一批的各个试验均匀地安排在试验区间内进行的方法.设做 $2n$ 个试验(n 为任意正整数),则将试验区间均分为 $2n+1$ 份,在 $2n$ 个分点 $x_1,x_2,\cdots,x_i,\cdots,x_{2n}$上做 $2n$ 个试验.若 x_i 处最好,则保留(x_{i-1},x_{i+1})部分,丢掉其余部分.将

留下部分均分为 $2n+2$ 份，在未做过试验的 $2n$ 个分点上再做试验，即将 $2n$ 个试验均匀地安排在好点的两旁，如此继续试验下去，便能找到最优点. 用这个方法，第一批试验后区间缩短 $2/(2n+1)$ 倍，以后每批试验后均分为 $n+1$ 份，在 n 个分点 $x_1, x_2, \cdots, x_i, \cdots, x_n$ 上做 n 次试验，将所得结果在同一条件下进行试验分析. 若 x_i 处最好，则保留 (x_{i-1}, x_{i+1}) 部分，丢掉其余所有部分. 然后，仍将 (x_{i-1}, x_{i+1}) 均分为 $n+1$ 份，再按上述方法处理，这样继续试验下去，以后每批试验后区间都缩短为前次留下的 $1/(n+1)$ 倍.

比例分割分批试验法(proportion cutting block search)　一种分批试验法. 是把第一批试验点按比例地安排在试验区间内分批做试验的方法. 对于试验区间[0,1]，若每批做四个试验，第一批在

$$\frac{5}{17}, \frac{6}{17}, \frac{11}{17}, \frac{12}{17}$$

四个点上进行；第二批将留下好点所在的线段六等分(共有 5 个分点)，在没有做过试验的四个分点上进行试验；以下每批四个试验点也总是在上次留下的好点的两侧均匀地安排着，如此继续下去. 第一批试验后范围缩短为 6/17，以后每批试验后都缩短为前次留下的 1/3. 比例分割分批试验法虽比均分分批试验法好，但由于比例分割分批试验法的试验点很接近，若试验效果差别不显著，则不好鉴别. 这种方法适用于当因素变动较小，而质量都有显著变化的情况.

时延试验法(time delay search)一种分批试验法. 是试验结果要"时延"一个周期，即隔一批才知道结果，或者两个周期，即隔两批才知道结果的方法.

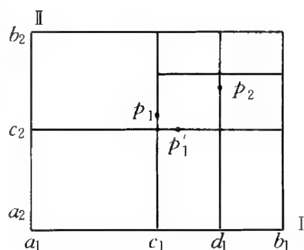
单因素爬山法(univariate climbing method)　亦称逐步提高法. 一种单因素优选法. 此方法是根据经验或估计(在成批生产中可以用原来生产的点)先找一个起点 A，在 A 点做试验后，向该因素的减少方向找一点 B 做试验. 如果好，就继续减少；如果不好，就往增加方向找一点 C 做试验. 如果 C 点好就继续增加，这样一步一步地提高. 如果爬到某点 M，再增加时反而坏了，则 M 就是该因素的最好点. 这种方法适用于某些可变因素不允许大幅度调整的情况，还适用于当某些可变因素要调到某点时，必须经过由小到大或由大到小的连续过程的情况. 爬山法的效果快慢不仅与起点选择有关，而且也与每步间隔大小的选择有关. 在实践中采用"两头小，中间大"的办法，即先在各个方向上用小步试探，找出有利于寻找目标的方向，当方向确定后，再根据具体情况跨大步，到快接近最好点时改为小步. 如果大步跨过最好点，这时可退回一步，在这一步内改用小步进行.

逐步提高法(step by step heighten method)　即"单因素爬山法".

多因素优选法(multidimensional optimum seeking method)　一种优选方法. 指在安排试验时，考虑两个或两个以上对目标最有影响的因素，进行合理安排，找到最优点或近似最优点，以期达到满意的试验结果的方法. 若是两个因素的，称为双因素优选法. 若是三个因素的，称为三因素优选法. 双因素优选法是迅速地找到二元函数 $f(x,y)$ 的最大值(或最小值)及其对应的最大点(或最小点)的方法，其几何解释是：要找出曲面 $z=f(x,y)$ 的最高峰(或最低谷). 如果把曲面 $z=f(x,y)$ 看做一座山的表面，设顶峰只有一个，把试验区域中 $f(x,y)$ 取同一值的曲线称为等高线，等高线一圈套一圈，越高越在里边. 所以，双因素优选法就是通过试验和比较的方法来寻找比较靠里边的等高线，直至找到最里边的最高点为止.

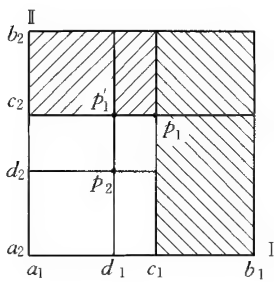
降维法(method of reduction dimensions)　一类优选方法. 是把一个多因素问题转化成一个较少因素(降低问题的维数)问题，而且较容易进行合理安排，找到最优点或近似最优点，以期达到满意的试验结果的方法. 纵横对折法、从好点出发法、平行线法等都是降维法.

纵横对折法(vertical and horizontal halver method)　一种降维法. 方法是：先固定一个因素于试验区域的中点，用单因素优选法对另一个因素进行选优；然后，固定这一个因素于试验区域的中点，再优选另一个因素；将两个结果进行比较，沿着"坏"点所在的线，丢去不包括好点所在的半个平面，这样继续下去，不断地将试验区域缩小，直至找到最优点. 设横坐标(x_1)表示第一个因素 Ⅰ，纵坐标(x_2)表示第二个因素 Ⅱ. Ⅰ 的试验区间为 $[a_1, b_1]$，Ⅱ 的试验区间为 $[a_2, b_2]$(如图). 先固定 Ⅰ 在试验区间的中点 c_1，即 $(b_1+a_1)/2$ 处，对 Ⅱ 进行单因素选优，设得到好点为 p_1. 然后固定 Ⅱ 在中点 c_2，即 $(b_2+a_2)/2$ 处，对 Ⅰ 进行单因素选优，得到好点 p_1'. 若 p_1' 比 p_1 好，则丢去平面：$a_1 \leqslant x_1 \leqslant c_1, a_2 \leqslant x_2 \leqslant b_2$ 部分，再在 Ⅰ 的新区间 $(c_1, b_1]$ 的中点 d_1，用单因素优选法优选因素 Ⅱ. 若好点为 p_2，而且 p_2 比 p_1' 好，则丢去平面 $c_1 < x_1 \leqslant b_1, a_2 \leqslant x_2 \leqslant c_2$ 部分. 如此继续下去，直至找到满意的结果为止.



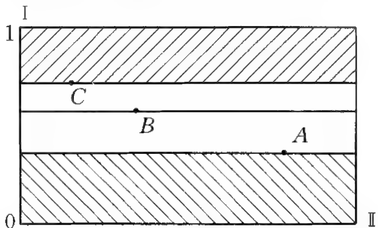
从好点出发法(method of start from good point)　一种降维法. 是对某一因素进行优选试验时，另一因素都固定在上次试验结果的好点上(除第一次外)的方法. 先固定因素 Ⅰ(x_1)于原生产点(或

0.618 点）c_1（如图所示），用单因素优选法优选因素Ⅱ（x_2），设得到好点为 $p_1(c_1,c_2)$. 然后，把因素Ⅱ固定在 c_2，用单因素优选法优选因素Ⅰ，得到好点 $p'_1(d_1,c_2)$. 去掉 p_1 右边的半面，试验区域缩为 $a_1 \leqslant x_1 < c_1, a_2 \leqslant x_2 \leqslant b_2$，再将因素Ⅰ固定在 d_1. 优选因素Ⅱ，设得到好点 $p_2(d_1,d_2)$，则去掉 p'_1 以上部分，试验范围为 $a_1 \leqslant x_1 < c_1, a_2 \leqslant x_2 < c_2$. 再将因素Ⅱ固定在 d_2，用单因素优选法在 $[a_1,c_1)$ 范围内优选因素Ⅰ，这样继续下去，就能找到所需要的最好点.



因素轮换法(elements rotate method) 即“从好点出发法”.

平行线法(parallel lines method) 一种降维法. 指每次试验都在相互平行的直线上做的方法. 平行线法是在实践中遇到一个因素不易改变的情况下应用的. 设影响目标的因素有Ⅰ和Ⅱ两个，因素Ⅰ难以调整，则先把难以调整的因素Ⅰ固定在0.382处（如图所示），用单因素优选法找出因素Ⅱ的最好点 A. 再将因素Ⅰ固定在0.618处，用单因素优选法找出因素Ⅱ的最好点 B. 比较 B 与 A，若 B 比 A 好，则丢掉 A 以下部分. 然后，按0.618法找出因素Ⅰ的第三点0.764. 第三次试验时，将因素Ⅰ固定在0.764处，用单因素优选法选出因素Ⅱ的最好点 C. 比较 C 与 B，若仍然 B 好，则丢去0.764以上部分. 如此继续做下去，直至找到满意的结果为止. 若三条平行线上的三个较好点在一条直线上，则以后的试验就不用平行线法做，而是在三点的连线上用单因素优选法找出最好点.



双因素爬山法(double factor climbing method) 一种双因素优选法. 该方法是先用单因素爬山法找到第一个因素最好点以后，以此点为起点，再对第二个因素应用同样方法找最好点. 若达到某点后，不论改变哪个因素，其结果都比这点差，则此点就是最优点. 或者以起点为中心，向四周搜索，找出有利于寻找目标的方向，在这个方向上跨一步，再搜索. 如此，边搜索边前进，直至找到最优点为止.

三角形对影法(contract shadow method of triangle) 亦称翻筋斗法. 一种双因素优选法. 该方法是先在等边三角形的三个顶点 A,B,C 处做三个试验，比较其结果，若 C 点最差，则在 C 的对称点 D 处做下一次试验. 若 A,B,D 三点中 B 点处最差，则下一次试验就在 B 的对称点处做. 如此每次丢去结果最差的一点，代之以该点的对称点，最后找出最优点. 三角形对影法也是一种双因素爬山法，是单纯形调优法的特例(参见“单纯形调优法”).

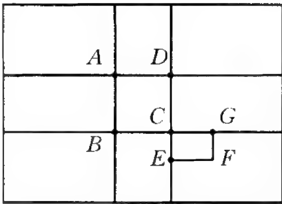
步长和方向双加速法(doubly speed up the method of step size anddirection) 亦称罗森布罗克法. 一种双因素优选法. 该方法是旋转原来的坐标轴，使得新的坐标轴方向逐步与山脊的方向吻合，消去各因素间的相互作用部分. 当山脊比较直时，用此法能较快地找到最高峰. 此法的基本思路仍旧是随着山脊方向向上爬. 1960年，罗森布罗克(Rosenbrock, H. H.)提出此法.

罗森布罗克法(Rosenbrock method) 即“步长和方向双加速法”.

陡度法(abrupt method) 一种爬山法(参见“单因素爬山法”). 先以起点为中心，在其四周做四个试验，然后分析其结果，从中找出最陡方向，在这个方向上继续试验. 用此法只作一次陡度试验所得的最好点不一定是最优点. 因此，可用上面同样的方法再做陡度试验，直至找到最优点或满意的结果为止.

混合法(mixed method) 一种实用的优选法. 它指根据具体问题，灵活地将几种优选方法结合起来应用的技巧. 如先用陡度法得到最优点或近似最优点的位置后，再改用其他方法，直至找到最优点或满意的结果为止.

矩形法(square method) 一种双因素优选法. 它分别通过两个因素试验范围的0.618点和0.382点，分别作两坐标轴的垂线，得 A,B,C,D 四个交点(如图所示)，在这四个点处做第一批试验. 若试验结果 C 点好，则第二批试验在 E,F,G 三点做. E,F,G,C 是新矩形的四个顶点，横向步长 CG、纵向步长 CE 可根据问题情况适当选定. 如此继续试验下去，直至找到满意的结果为止. 若经过三次试验后，好点仍旧不变，则需缩短步长，找出新的试验点，用新的步长按照上述方法试验，直至找到满意的结果. 如果结果还不满意，可在原好点的两侧另做矩形继续试验下去. 但需注意，做完一次试验后不能丢去一块，因为这样可能把最好点丢掉.



平分平面法(divide equally plane method) 一种三因素优选法. 它在三个平分平面：$X=1/2, 0 \leqslant Y \leqslant 1, 0 \leqslant Z \leqslant 1$；$0 \leqslant X \leqslant 1, Y=1/2, 0 \leqslant Z \leqslant 1$；$0 \leqslant X \leqslant 1, 0 \leqslant Y \leqslant 1, Z=1/2$ 上用双因素优选法找出最好

点，其中 X,Y,Z 分别表示三个因素. $0\leqslant X\leqslant 1,0\leqslant Y\leqslant 1,0\leqslant Z\leqslant 1$ 表示立方体. 设$(1/2,Y_1,Z_1)$为较好点，且 $0\leqslant Y_1\leqslant 1/2,0\leqslant Z_1\leqslant 1$，则切掉立方体的 3/4，留下好点所在的立方体 $0\leqslant X\leqslant 1,0\leqslant Y\leqslant 1/2,0\leqslant Z\leqslant 1/2$. 在此立方体中用同样方法继续进行优选.

平行平面法(parallel plane method)　一种优选方法. 首先任取一个因素，在两个平行平面上用双因素优选法找出最好的点，切掉其余部分，留下好点所在的那个部分，然后继续进行优选.

随机试验法(stochastic test method)　一类优选方法. 它是在试验范围内随机地选择试验点，从中选择优化途径的一类方法. 如下的随机方向法便是此类方法之一：它从某点 P_0 出发，随机地选取方向和步长，得到新点 P_1；若 P_1 比 P_0 好，则沿此方向使原步长加长一倍到 P_2；若 P_2 比 P_1 好，继续如此试验；若 P_2 比 P_1 的试验结果差，则缩回到 P_1 点，再随机地选取方向和步长. 如此继续下去，直至找到满意的结果为止.

排　队　论

排队论(queueing theory)　亦称随机服务系统理论. 运筹学的一个重要分支. 是研究排队现象随机规律的一门学科. 20 世纪初，丹麦工程师埃尔朗(Erlang, A. K.)等人对电话通话中的排队现象进行了研究，发表了"概率论与电话通话"的论文，最早把排队论用于电话话务理论中. 后来，排队论逐渐应用到机器管理、交通运输、纺织、水库设计和可靠性理论等，近年来扩展到电信、计算机设计等领域. 排队论的主要内容有：

1. 研究各种排队系统的概率规律性，如队长分布、等待时间分布和忙期分布等. 这类问题为性态问题.

2. 研究最优设计(即静态最优)和排队系统的最优运营(即动态最优). 这类问题为最优化问题.

3. 根据排队理论，研究分析一个给定的排队系统符合那种模型. 这类问题为排队系统的统计推断问题.

排队系统由输入过程(或称到达流)、排队规则和服务机构三个基本要素组成.

排队系统的基本要素(fundamental elements of queueing system)　见"排队论".

随机服务系统理论(random service system)　即"排队论".

输入过程(input process)　排队论的基本概念之一. 指顾客到达排队系统的过程(情况). 顾客总体(称为顾客源)可能是有限的，也可能是无限的；顾客到来的方式可能是单个的，也可能是成批的；顾客相继到达的间隔时间可以是确定型的，也可以是随机型的；顾客的到达可以是相互独立的；输入过程可以是平稳的，或称对时间是齐次的(参见"平稳输入").

到达流(arrival stream)　即"输入过程".

平稳输入(stationary input)　亦称时间是齐次的输入. 一种输入过程. 是指在时间区间$[a,a+t)$内有 n 个顾客到达的概率与 a 无关，而仅与 t 及 n 有关的输入，如 k 阶埃尔朗输入、一般独立输入、成批输入等. 否则，便是非平稳输入，如电话系统、呼叫流就是非平稳的，而且具有明显的周期性. 除平稳输入和非平稳输入外，还有其他输入，如有调度员的输入(当有多个平行的排队队列时，顾客到达由调度员按排加入哪个队)，还有半马氏输入.

顾客(customer)　排队论术语. 指在排队系统中，需要进行某种服务的对象. 顾客可以是人，也可以是物等.

服务台(server)　亦称服务员. 排队论术语. 指在排队系统中，对顾客进行服务的人或设施. 服务台可以是一个，也可以是多个.

排队规则(queue discipline)　排队论的基本概念之一. 指顾客到达时，若所有服务台都正在被占用，则顾客可以随即离去，也可以排队等候的规则. 随即离去的称为损失制或即时制；排队等候的称为等待制. 对于等待制，为顾客服务的次序可以采用先到先服务、后到先服务、随机服务、有优先权的服务等规则. 队列可以排在具体的处所，也可以是抽象的. 有的系统规定容量有限，有的系统规定容量无限. 队列的数目可以是单列，也可以是多列. 在多列中，各列间的顾客有的可以互相转移，有的不能；有的可以中途退出，有的不能.

损失制(loss system)　亦称即时制. 见"排队规则".

损失制服务系统(loss service system)　亦称消失制服务系统. 一种排队系统. 指顾客到达时，系统中所有服务台都被占用，此时顾客随即离去.

消失制服务系统(loss service system)　即"损失制服务系统".

混合制服务系统(mixed service system)　一种排队系统. 是等待制服务系统与损失制服务系统相混合的系统. 这类系统有两种情形：一是顾客等候的时间有限，在此有限时间内等待服务，超过此有限时间随即离去；另一是系统内排队等候的容量有限，能容纳时则等候，否则离去.

等待时间(waiting time)　排队论术语. 指顾客进入服务系统后到接受服务前排队等待的时间. 以 w_i 表示第 i 名顾客的等待时间，有如下关系：

$$w_{i+1}=\begin{cases}w_i+s_i-t_i & (w_i+s_i-t_i>0),\\ 0 & (w_i+s_i-t_i<0),\end{cases}$$

其中 t_i 表示到达间隔时间，即 $t_i=\tau_{i+1}-\tau_i$，τ_i 表示第 i 号顾客到达的时刻，s_i 表示服务时间.

等待制(waiting system)　见“排队规则”.

等待制服务系统(waiting service system)　一种排队系统. 指顾客到达时，若系统中所有服务台都被占用，此时顾客排队等候，直至接受服务后再离去.

服务机构(service mechanism)　排队论术语. 指服务方式的描述. 有以下几种情况：

1. 服务台可以有一个或多个.

2. 在有多个服务台时，可以平行排列(并列)，也可以前后排列(串列)，也可以混合排列.

3. 服务可以对单个顾客进行，也可以对成批顾客进行.

4. 服务时间分为确定型和随机型，随机型的要已知其经验分布或理论(概率)分布.

排队模型记法(notation of queueing model)　排队模型的符号表示法. 用一斜线“/”分开，如 $X/Y/Z/C$，第一个符号 X 表示相继到达间隔时间的分布，第二个符号 Y 表示服务时间的分布，第三个符号 Z 表示服务台的数目，第四个符号 C 表示排队系统中允许的最大顾客数(容量). $Z\leqslant C\leqslant+\infty$，当 $C=Z$ 时，表示损失制；当 $Z<C$(有限制)时，表示混合制；当 $C=+\infty$时，表示等待制. 1953 年，肯德尔(Kendall, D. G.)提出了上述符号形式，这就是排队模型的记法，或称为经典排队模型(系统)的符号表示法.

分布符号(distribution sign)　排队论中的一些符号. 指用来表示顾客到达时间间隔或服务时间分布的符号. M 表示负指数分布(M 是 Markov 的字头)，D 表示确定型，E_k 表示 k 阶埃尔朗分布，GI 表示一般相互独立的随机分布，G 表示一般随机分布. 例如，$M/M/1$ 表示相继到达时间间隔为负指数分布、服务时间为负指数分布、单服务台的模型；$D/M/C$ 表示确定的到达间隔、服务时间为负指数分布、C 个平行服务台(顾客是一队)的模型；$GI/G/1$ 表示有一般相互独立的随机到达、一般随机服务时间和单服务台的模型.

排队模型的基本数量指标(basic quantitative index of queueing model)　排队论中几个用数量表示的指标. 包括队长、队列长、逗留时间、等待时间和忙期等. 队长是系统中的顾客数，其期望值记为 L_s. 队列长是系统中排队等待服务的顾客数，其期望值记为 L_q，系统中顾客数等于在队列中等待服务的顾客数与正被服务的顾客数之和，L_q(或 L_s)越大，服务率越低. 逗留时间是一个顾客在系统中的停留时间，其期望值记为 W_s，逗留时间等于等待时间与服务时间之和.

队长(team length)　见“排队模型的基本数量指标”.

队列长(queue length)　亦称排队长. 见“排队模型的基本数量指标”.

逗留时间(sojourn time)　见“排队模型的基本数量指标”.

系统状态(state of system)　排队论的基本概念之一. 指系统中的顾客数，即队长. 若系统中有 n 个顾客，则系统的状态是 n. 随着情况的不同，它的可能值是：

1. 队长没有限制时，即 $n=0,1,2,\cdots$.

2. 队长有限制，最大数为 N 时，即

$$n=0,1,2,\cdots,N.$$

3. 即时制，服务台个数是 C 时，即

$$n=0,1,2,\cdots,C.$$

在时刻 t，系统状态为 n 的概率称为状态概率，用 $P_n(t)$表示. 求状态概率的方法是建立含 $P_n(t)$的关系式，这个关系式一般是微分差分方程(关于 t 的微分方程，关于 n 的差分方程). 含 $P_n(t)$方程的解，称为瞬态解.

状态概率(state probability)　见“系统状态”.

稳态(steady state)　亦称统计平衡状态. 排队论的基本概念之一. 指排队系统的概率统计特性不随时间而改变的状态. 对于一个实际排队系统，是指当系统运行充分长的时间后，初始($t=0$)出发状态的概率分布($P_n(0),n\geqslant 0$)的影响逐渐消失，而且系统的状态概率分布 $P_n(t)$趋近于一常数 P_n. 稳态概率常用

$$\lim_{t\to\infty} P_n(t) = P_n$$

表示. 求稳态概率 P_n 时，只需令导数$P'_n(t)=0$ 即可.

统计平衡状态的解(solution of statistical equilibrium state)　即稳态解. 这种解构成一个状态概率分布：$\{P_n(t),n\geqslant 0\}$，亦称为稳定状态概率.

泊松流(Poisson process)　亦称最简单流. 排队论的基本概念之一. 记 $N(t)$表示在时间区间$(0,t]$($t>0$)内到达的顾客数(约定 $N(0)=0$)，在时间区间$(t_1,t_2]$($t_2>t_1$)内有 n 个顾客到达的概率为

$$P_n(t_1,t_2)=P\{N(t_2)-N(t_1)=n\}\quad(n\geqslant 0).$$

当$\{N(t),t\geqslant 0\}$满足下面三个条件时，则称此输入过程$\{N(t),t\geqslant 0\}$为泊松流. 这三个条件是：

1. 无后效性. 在不相重叠的时间区间内，顾客到达数是相互独立的.

2. 平稳性. 在时间区间$[a,a+t)$内有 n 个顾客到达的概率 $P_n(a,a+t)$与 a 无关，而仅与 t 及 n 有关，即 $P_n(a,a+t)=P_n(0,t)$.

3. 普通性. 对于充分小的 Δt，在时间区间$[t,t+$

$\Delta t)$内有两个或两个以上顾客到达的概率极小，即有

$$P_1(t,t+\Delta t)=\lambda\Delta t+o(\Delta t),$$

$$\sum_{n=2}^{\infty}P_n(t,t+\Delta t)=o(\Delta t),$$

其中 $\lambda>0$ 是常数，$o(\Delta t)$是 Δt 的高阶无穷小.

当输入过程为泊松流时，在长为 t 的时间区间内到达 n 个顾客的概率遵从泊松分布，即

$$P_n(a,a+t)=P_n(0,t)=\frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

$$(t>0,n=0,1,2,\cdots),$$

而在单位长的时间区间内到达的平均顾客数为 λ，即 $E(N(a+1)-N(a))=\lambda$.

负指数分布(negative exponential distribution) 一种概率分布. 设随机变量 T 的概率分布为

$$F_T(t)=P\{T\leqslant t\}=1-e^{-\lambda t}\quad(t\geqslant 0),$$

式中 $\lambda>0$ 为常数，则称 T 服从参数是 λ 的负指数分布. T 的数学期望和方差分别是

$$E[T]=\frac{1}{\lambda}\text{ 和 }\mathrm{var}[T]=\frac{1}{\lambda^2}.$$

负指数分布具有无记忆性. 若把 T 解释为电子元件的工作寿命，而且服从负指数分布，则不论它已经工作多长时间，它的剩余寿命的条件分布与原分布相同，不受已有年龄的影响.

顾客到达的n 概率分布(n probability distribution of customer arrive) 一种概率分布. 在时刻 t，系统状态为 n 的概率分布(参见“系统状态”)为

$$P_n(t)=\frac{(\lambda t)^n}{n!}e^{-\lambda t}\quad(t>0,n=0,1,2,\cdots).$$

由泊松流的条件 2，有 $P_n(t)=P_n(0,t)$，它表示长为 t 的时间区间内到达 n 个顾客的概率，且时间由 0 算起. 由泊松流的条件 2 和 3，在$[t,t+\Delta t)$区间内没有顾客到达的概率为

$$P_0(t+\Delta t)=1-\lambda\Delta t+o(\Delta t).$$

随机变量$\{N(t)=N(S+t)-N(S)\}$服从泊松分布，它的数学期望和方差分别是 $E[N(t)]=\lambda t$ 和 $\mathrm{var}[N(t)]=\lambda t$. 下表列出在三个区间上三种情况下顾客到达的个数和概率.

$[0,t)$		$[t,t+\Delta t)$		$[0,t+\Delta t)$	
个数	概 率	个数	概 率	个数	概 率
n	$P_n(t)$	0	$1-\lambda\Delta t+o(\Delta t)$	n	$P_n(t)(1-\lambda\Delta t+o(\Delta t))$
$n-1$	$P_{n-1}(t)$	1	$\lambda\Delta t$	n	$P_{n-1}(t)\lambda\Delta t$
$n-2$	$P_{n-2}(t)$	2	$o(\Delta t)$	n	$o(\Delta t)$
$n-3$	$P_{n-3}(t)$	3		n	
⋮	⋮	⋮		⋮	
0	$P_0(t)$	n		n	

服务时间的分布(distribution of service time) 一种概率分布. 指服务台对每一顾客服务时间的概率分布. 例如，确定型的有定长分布；随机型的有负指数分布、埃尔朗分布、超指数分布等. 设服务时间服从负指数分布，它的分布函数和概率密度分别是

$$F_T(t)=1-e^{-\lambda t}\text{ 和 }f_T(t)=\lambda e^{-\lambda t},$$

其中 λ 表示单位时间能被服务完的顾客数(期望值)，称为平均服务率，而 $1/\lambda$ 表示一个顾客的平均服务时间，即期望值.

埃尔朗分布(Erlang distribution) 一种概率分布. 其分布密度为

$$b_k(t)=\frac{k\mu(k\mu t)^{k-1}}{(k-1)!}e^{-k\mu t}\quad(t>0).$$

令

$$T=v_1+v_2+\cdots+v_k,$$

其中 $v_1,v_2,\cdots,v_k$ 是 k 个相互独立的随机变量，服从相同参数 $k\mu$ 的负指数分布，则 T 服从 k 阶埃尔朗分布. T 和 T^n 的数学期望以及 T 的方差依次是

$$E[T]=\frac{1}{\mu},$$

$$E[T^n]=\frac{k(k+1)\cdots(k+n-1)}{(k\mu)^n}\quad(n\geqslant 1),$$

$$\mathrm{var}[T]=\frac{1}{k\mu^2}\quad.$$

当 $k=1$ 时，埃尔朗分布化为负指数分布；当 k 增大到不小于 30 时，埃尔朗分布近似于正态分布；当 $k\to\infty$时，因 $\mathrm{var}[T]\to 0$，故埃尔朗分布可以化为确定型分布；一般 k 阶埃尔朗分布可以看成随机与确定的中间型. 埃尔朗分布本身虽不具有无记忆性，但仍可利用负指数分布的无记忆性.

超指数分布(hyperexponential distribution) 一种概率分布. 有 k 个平行的服务台，服务时间均服从负指数分布，平均服务时间分别为 $1/\mu_i(i=1,2,\cdots,k)$，一个顾客到达后以概率 α_i 选取第 i 个服务台，但直到正在接受服务的顾客服务完成之前，不允许新的顾客在别的服务台处接受服务，这样顾客的服务时间分布就服从 k 阶超指数分布.

话务强度(traffic intensity) 排队论术语. 指单位时间内顾客到达的平均数与能被服务的平均数之比，即 $\rho=\lambda/\mu$. 称 ρ 为话务强度，或称为服务强度.

运行指标(operational index) 排队论中的一些数量指标. 如系统中队长的期望值 L_s，队列中队列长的期望值 L_q，系统中顾客逗留时间的期望值 W_s 和在队列中顾客等待时间的期望值 W_q.

$M/M/1$ 模型($M/M/1$ model) 一种排队模型. 指输入过程服从泊松流，服务时间服从负指数分布，单服务台的排队模型. 该模型分为标准的 $M/$

$M/1$ 模型、系统的容量有限制，以及顾客源为有限三种情形.

标准 $M/M/1$ 模型(standard $M/M/1$ model)　一种排队模型. 指队长没有限制，先到先服务，到达间隔和服务时间是相互独立的排队模型.

标准 $M/M/1$ 模型的队长期望值为

$$L_s = \sum_{n=0}^{\infty} nP_n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}\ ;$$

队列长期望值为

$$L_q = \sum_{n=1}^{\infty}(n-1)P_n = L_s - \rho = \frac{\rho^2}{1-\rho} = \frac{\rho\lambda}{\mu-\lambda}\ ;$$

逗留时间的期望值为

$$W_s = E[W] = \frac{1}{\mu-\lambda}\ ;$$

等待时间的期望值为

$$W_q = W_s - \frac{1}{\mu} = \frac{\rho}{\mu-\lambda}\ .$$

标准 $M/M/1$ 模型的稳态解 $P_n=(\lambda/\mu)^n P_0$ (P_n 表示 $P_n(t)$ 与 t 无关)，由微分差分方程

$$\begin{cases}\lambda P_{n-1} + \mu P_{n+1} - (\lambda+\mu)P_n = 0 \quad (n=1,2,\cdots), \\ -\lambda P_0 + \mu P_1 = 0\end{cases}$$

求出. 令 $\rho=-\lambda/\mu<1$，且

$$\sum_{n=0}^{\infty} P_n = 1,$$

则

$$P_0\sum_{n=0}^{\infty}\rho^n = P_0\frac{1}{1-\rho} = 1,$$

从而得到稳态解为

$$P_0 = 1-\rho\ ,\ P_n = (1-\rho)\rho^n \quad (n=1,2,\cdots).$$

队长期望值(expected value of team length)　见“标准 $M/M/1$ 模型”.

队列长期望值(expected value of queueing length)　见“标准 $M/M/1$ 模型”.

逗留时间的期望值(expected value of sojourn time)　见“标准 $M/M/1$ 模型”.

等待时间的期望值(expected value of waiting time)　见“标准 $M/M/1$ 模型”.

标准 $M/M/1$ 模型的稳态解(steady state solution of standard $M/M/1$ model)　见“标准 $M/M/1$ 模型”.

生灭过程(birth and death process)　一种随机过程. 指系统状态随时间变化的一种过程. 设有某个系统，具有状态集 $S=\{0,1,2,\cdots\}$；$N(t)$ 表示系统在时刻 $t(\geqslant 0)$ 所处的状态；在时刻 t 系统处于状态 n 的条件下，若再经过长为 Δt 的时间段，系统的状态 $N(t+\Delta t)$ 满足下列三个条件，则称为一个生灭过程. 这三个条件是：

1. 转移到状态 $n+1(0\leqslant n<+\infty)$ 的概率为 $\lambda_n\Delta t+o(\Delta t)$.

2. 转移到状态 $n-1(1\leqslant n<+\infty)$ 的概率为 $\mu_n\Delta t+o(\Delta t)$.

3. 转移到状态 $A\equiv S-\{n-1,n,n+1\}$ 的总概率为 $o(\Delta t)$，其中 $\lambda_n>0, \mu_n>0$ 为固定常数，且与 t 无关.

若 S 仅包含有限个元素：$S=\{0,1,2,\cdots,k\}$，也满足以上三个条件，但要把前两个条件中 n 的变化范围改为：

1′. $0\leqslant n<k$；

2′. $1\leqslant n\leqslant k$；

则称为有限状态生灭过程. 对于标准 $M/M/1$ 模型，系统状态 $N(t)$ 为一生灭过程. 此生灭过程的微分差分方程是

$$\frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) + \mu P_{n+1}(t) - (\lambda+\mu)P_n(t) \quad (n\geqslant 1),$$

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t),$$

式中 $P_n(t)$ 表示系统中有 n 个顾客的概率，到达规律服从参数为 λ 的泊松流，服务时间服从参数为 μ 的负指数分布.

忙期(busy period)　排队论术语. 指排队系统中，由顾客数为 0 的状态(即空闲状态)后的第一个顾客到达且立即被服务的时刻开始，到顾客数再度为 0(再度空闲)为止的时间长度. 在此期间，系统处在繁忙状态，即有顾客等待或正被服务. 设忙期的状态为 $N(t)$，则 $N(t)$ 由 0 以后第一次变为 1 的时刻忙期开始，其后 $N(t)$ 又首次变为 0 时忙期就结束. 即 $t=0$ 为忙期的起点，此时 $N(0)=1$，而 $N(t)$ 首次变为 0 的时刻为忙期的结束.

闲期(idle period)　排队论术语. 指排队系统中，从顾客数等于 0 的时刻开始，到重新有顾客到达时刻为止的时间长度. 系统处在空闲状态的概率 $P_0=1-\rho$. 闲期的期望值 $\bar{I}=1/\lambda$，这里的条件是到达时间间隔服从负指数分布.

$M/M/1(N)$ 模型($M/M/1(N)$ model)　一种排队模型. 是系统容量为 N 的 $M/M/1$ 模型. 当 $N=1$ 时为即时服务制的情形；当 $N\to\infty$ 时为容量无限制的情形.

设简单流输入的参数为 λ，指数服务分布的参数为 μ，单个服务台，容量为 N，则系统中统计平衡状态的解称为 $M/M/1(N)$ 模型的稳态解. 它是

$$P_0=\frac{1-\rho}{1-\rho^{N+1}} \quad (\rho\neq 1),$$

$$P_n=\frac{1-\rho}{1-\rho^{N+1}}\rho^n\quad(\rho\neq 1;n\leqslant N).$$

若 $\rho=1(\lambda=\mu)$，则

$$P_0=P_1=\cdots=\frac{1}{N+1}.$$

这时，队长期望值为

$$L_s=\sum_{n=0}^{N}nP_n=\frac{\rho}{1-\rho}-\frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}\quad(\rho\neq 1);$$

队列长期望值为

$$L_q=\sum_{n=1}^{N}(n-1)P_n=L_s-(1-P_0);$$

有效到达率为

$$\lambda_e=\mu(1-P_0)=\lambda(1-P_N);$$

逗留时间的期望值为

$$W_s=\frac{L_s}{\mu(1-P_0)}=\frac{L_q}{\lambda(1-P_N)}+\frac{1}{\mu};$$

等待时间的期望值为

$$W_q=W_s-\frac{1}{\mu}.$$

***M/M/1(N)*模型的稳态解**(steady state solution of *M/M/1(N)* model)　见“*M/M/1(N)*模型”.

***M/M/1(m)*模型**(*M/M/1(m)* model)　一种排队模型. 是顾客源 m 为有限的排队模型，其他特征与 $M/M/1$ 模型相同. 此模型的稳态解为

$$P_0=\frac{1}{\sum_{i=0}^{m}\frac{m!}{(m-i)!}\left(\frac{\lambda}{\mu}\right)^i},$$

$$P_n=\frac{m!}{(m-n)!}\left(\frac{\lambda}{\mu}\right)^nP_n\quad(1\leqslant n\leqslant m).$$

其队长期望值为

$$L_s=m-\frac{\mu}{\lambda}(1-P_0).$$

在机器故障问题中，L_s 表示平均故障台数，而

$$m-L_s=\frac{\mu}{\lambda}(1-P_0)$$

表示正常运转的平均台数. 队列长期望值为

$$L_q=m-\frac{(\lambda+\mu)(1-P_0)}{\lambda}=L_s-(1-P_0);$$

逗留时间期望值为

$$W_s=\frac{m}{\mu(1-P_0)}-\frac{1}{\lambda};$$

等待时间期望值为

$$W_q=W_s-\frac{1}{\mu}.$$

***M/M/C* 模型**(*M/M/C* model)　一种排队模型. 指单队、并列的多服务台的排队模型，C 表示服务台数. 该模型分为标准 $M/M/C$ 模型、系统容量有限制(N)，以及有限顾客(m)三种情形. $M/M/C$ 模型的稳态解为

$$P_0=\left[\sum_{k=0}^{C-1}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k+\frac{1}{C!}\frac{1}{1-\rho}\left(\frac{\lambda}{\mu}\right)^C\right]^{-1},$$

$$P_n=\begin{cases}\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^nP_0 & (n<C),\\ \frac{1}{C!C^{n-C}}\left(\frac{\lambda}{\mu}\right)^nP_0 & (n\geqslant C).\end{cases}$$

其平均队长为

$$L_s=L_q+C\rho;$$

队列长期望值为

$$L_q=\sum_{n=C+1}^{\infty}(n-C)P_n=\frac{(C\rho)^C\rho}{C!(1-\rho)^2}P_C;$$

平均逗留时间为

$$W_s=W_q+\frac{1}{\mu}=\frac{L_s}{\lambda};$$

平均等待时间为

$$W_q=\frac{L_q}{\lambda}.$$

标准 *M/M/C* 模型(standard *M/M/C* model)　一种排队模型. 它与标准 $M/M/1$ 模型所规定的各种特征相同，服务台工作相互独立，平均服务率 $\mu_1,\mu_2,\cdots,\mu_C$ 相同，且只当 $\lambda/C\mu<1$ 时不会排成无限的队列，其中 $C\mu$ 为整个服务机构的平均服务率. 服务强度 $\rho=\lambda/C\mu$，服务强度亦称为服务台的平均利用率.

***M/M/C(N)*模型**(*M/M/C(N)* model)　一种排队模型. 是其系统的容量最大限制为 $N(N\geqslant C)$，即当系统中顾客数 n 已达到 N 时，再来的顾客即被拒绝，其他条件与标准 $M/M/C$ 模型相同的排队模型. 其稳态解为

$$P_0=\frac{1}{\sum_{k=0}^{C}\frac{(C\rho)^k}{k!}+\frac{C^C}{C!}\frac{\rho(\rho^C-\rho^N)}{1-\rho}}\quad(\rho\neq 1),$$

$$P_n=\begin{cases}\frac{C(\rho)^n}{n!}P_0 & (0\leqslant n\leqslant C),\\ \frac{C^C}{C!}\rho^nP_0 & (C\leqslant n\leqslant N).\end{cases}$$

在上式中 $N=C$ 时，有

$$\begin{cases}P_0=\frac{1}{\sum_{k=0}^{C}\frac{(C\rho)^k}{k!}},\\ P_n=\frac{C(\rho)^n}{n!}P_0\quad(0\leqslant n\leqslant C)\end{cases}\quad\left(\rho=\frac{\lambda}{C\mu}\right).$$

运行指标：

$$L_q=\frac{P_0\rho(C\rho)^C}{C!\ (1-\rho)^2}[1-\rho^{N-C}-(N-C)\rho^{N-C}(1-\rho)];$$

$L_s=L_q+C\rho(1-P_N)$;

$W_q=\frac{L_q}{\lambda(1-P_N)}$;

$W_s=W_q+\frac{1}{\mu}$.

若式中 $N=C$,则:

$L_q=0$;

$$L_s=\sum_{n=1}^{C}nP_n=\frac{C\rho\sum_{n=0}^{C-1}\frac{(C\rho)^{n-1}}{n!}}{\sum_{n=0}^{C}\frac{(C\rho)^n}{n!}}$$

$=C\rho(1-P_C)$;

$W_q=0$;

$W_s=\frac{1}{\mu}$.

$M/M/C(m)$**模型**($M/M/C(m)$ model) 一种排队模型.指顾客源为有限数 m 的排队模型,且 $m>C$.有限源服务系统的典型例子是机器看管问题.设 m 表示机器台数,C 表示修理工的人数,顾客到达表示机器出了故障,每个顾客到达率 λ 表示每台机器每单位运转时间出故障的期望次数,顾客数 n 表示出故障的机器台数.$n\leqslant C$ 表示所有的故障机器都在被修理,且有 $(C-n)$ 个修理工人在空闲,$C<n\leqslant m$ 表示有 $(n-C)$ 台机器在停机等待修理,修理工人都处在繁忙状态.设这 C 个工人修理技术相同,修理时间都服从参数为 μ 的负指数分布,并假定故障的修复时间和正在生产的机器是否发生故障是相互独立的.其状态概率为

$$P_0=\frac{1}{m!}\Big/\left[\sum_{k=0}^{C}\frac{1}{k!(m-k)!}\left(\frac{C\rho}{m}\right)^k+\frac{C^C}{C!}\sum_{k=C+1}^{m}\frac{1}{(m-k)!}\left(\frac{\rho}{m}\right)^k\right],$$

式中 $\rho=\frac{m\lambda}{C\mu}$,

$$P_n=\begin{cases}\frac{m!}{(m-n)!n!}\left(\frac{\lambda}{\mu}\right)^nP_0 & (0\leqslant n\leqslant C),\\ \frac{m!}{(m-n)!C!C^{n-C}}\left(\frac{\lambda}{\mu}\right)^nP_0 & (C+1\leqslant n\leqslant m).\end{cases}$$

运行指标:

$$L_s=\sum_{n=1}^{m}nP_n=L_q+\frac{\lambda_e}{\mu}=L_q+\frac{\lambda}{\mu}(m-L_s);$$

$$L_q=\sum_{n=C+1}^{m}(n-C)P_n;$$

$$W_s=\frac{L_s}{\lambda_e};$$

$$W_q=\frac{L_q}{\lambda_e};$$

式中 $\lambda_e=\lambda(m-L_s)$,称为有效到达率,它等于每个顾客的到达率 λ 乘以在系统外机器的期望数.

$M/G/1$ **模型**($M/G/1$ model) 一种排队模型.指服务时间 T 的分布是一般的,但要求期望值 $E[T]$ 和方差 $\mathrm{var}[T]$ 都存在,其他条件和标准的 $M/M/1$模型相同的排队模型.其运行指标为

$$L_s=\rho+\frac{\rho^2+\lambda^2\mathrm{var}[T]}{2(1-\rho)},$$

式中 $\rho=\lambda E[T]<1$,称上式为 P-K 公式.只要知道 λ,$E[T]$和$\mathrm{var}[T]$,无论 T 是什么分布,都可以求出 L_s.然后,通过 $W_s=W_q+E[T]$,$L_s=\lambda W_s$ 和 $L_q=\lambda W_q$ 可以求 L_q,W_q 和 W_s.值得注意的是,要想改进上述各指标,因仅当 $\mathrm{var}[T]=0$ 时,随机性的波动才不影响 L_s,故除考虑期望值外,还可以用改变方差来考虑.对 $M/G/1/\infty$系统,肯德尔(Kendall,D. G.)首先使用了嵌入马氏链的分析方法.虽然该系统的队长过程$\{N(t),t\geqslant0\}$不具有马氏性,但在$\{N(t),t\geqslant0\}$中可嵌入一个离散参数的马氏链.利用马氏链的理论得到嵌入马氏链的极限概率分布,该分布就是原过程$\{N(t),t\geqslant0\}$当 $t\to\infty$时的平稳分布,顾客离去时刻是该过程的再生点.时齐的、非周期的可列状态马氏链$\{N_n,n\geqslant1\}$,其一步转移概率矩阵为

$$(P_{ij})=\begin{vmatrix}a_0 & a_1 & a_2 & a_3 & \cdots\\ a_0 & a_1 & a_2 & a_3 & \cdots\\ 0 & a_0 & a_1 & a_2 & \cdots\\ 0 & 0 & a_0 & a_1 & \cdots\\ \vdots & \vdots & \vdots & \vdots & \end{vmatrix}.$$

马氏链$\{N_n,n\geqslant1\}$是遍历的充分必要条件为 $\lambda/\mu=\rho<1$.当 $\rho<1$ 时,

$$\lim_{n\to\infty}P_{ij}^{(n)}=\pi_j>0\quad(i,j=0,1,2,\cdots),$$

且$\{\pi_j\}$为惟一的平稳分布;当 $\rho\geqslant1$ 时,

$$\lim_{n\to\infty}P_{ij}^{(n)}=0\quad(i,j=0,1,2,\cdots),$$

且不存在平稳分布,其中$(P_{ij}^{(n)})$为嵌入马氏链的 n 步转移矩阵.对任何正整数 m,有

$$\lim_{n\to\infty}P\{N_n\leqslant m\}=\begin{cases}\sum_{j=0}^{m}\pi_j>0 & (\rho<1),\\ 0 & (\rho\geqslant1).\end{cases}$$

当 $\rho\geqslant1$ 时,不论怎样大的 m,当 $n\to\infty$时,第 n 个顾客服务结束,离开时系统中留下的队长$\leqslant m$ 的概率总趋于 0.这说明队将愈来愈长,系统达不到统计平衡,即系统是不稳定的.若 $\rho=1$,则$\{N_n,n\geqslant1\}$为零常返链;若 $\rho>1$,则$\{N_n,n\geqslant1\}$为非常返链.对这两种情形均不存在平稳分布.当 $\rho=1$ 时,服务机构始终不空的概率为 0;而当 $\rho>1$ 时,服务机构始终不空的概率为正.当 $\rho<1$ 时,

$$\pi(z)=(1-\rho)\frac{(1-z)A(z)}{A(z)-z}$$

是马氏链$\{N_n,n\geqslant1\}$平稳状态分布$\{\pi_j\}$的母函数的表达式.$\pi_n=P_n$,$n\geqslant0$,π_n 就是在统计平衡下,任一时

刻系统有 n 个顾客的概率.

$M/D/1$ **模型**($M/D/1$ model) 一种排队模型. 指服务时间 T 是确定常数、单服务台的排队模型. 有

$$T=\frac{1}{\mu},\ \mathrm{var}[T]=0,\ L_s=\rho+\frac{\rho^2}{2(1-\rho)}.$$

$M/E_k/1$ **模型**($M/E_k/1$ model) 一种排队模型. 指顾客必须经过 k 个服务站,在每个服务站的服务时间 T_i 相互独立,服从相同的参数为 $k\mu$ 的负指数分布,而且

$$T=\sum_{i=1}^{k}T_i$$

服从 k 阶埃尔朗分布的排队模型. 这时有

$$E[T_i]=\frac{1}{k\mu},\ \mathrm{var}[T_i]=\frac{1}{k^2\mu^2},$$

$$E[T]=\frac{1}{\mu},\ \mathrm{var}[T]=\frac{1}{k\mu^2}.$$

系统主要运行指标为:

$$L_s=\rho+\frac{\rho^2+\dfrac{\lambda^2}{k\mu^2}}{2(1-\rho)}=\rho+\frac{(k+1)\rho^2}{2k(1-\rho)};$$

$$L_q=\frac{(k+1)\rho^2}{2k(1-\rho)};$$

$$W_s=\frac{L_s}{\lambda};$$

$$W_q=\frac{L_q}{\lambda}.$$

排队系统的最优化问题(optimization problem of queueing system) 一类最优化问题. 指排队系统设计的最优化问题. 分静态问题和系统控制最优化(动态问题)两类. 例如,为了提高服务水平,将会降低顾客的等待费用,增加服务机构的成本,这里最优化的目标之一是使二者费用之和为最小,以期达到这个目标为最优的服务水平. 另一个是使纯收入或利润目标(服务收入与服务成本之差)为最大.

$M/M/1$ **模型中最优服务率**(optimum service rate in $M/M/1$ model) 排队论术语. 指 $M/M/1$ 模型中,在给定的某最优化准则(如服务台的单位时间平均费用最小)下的一种数量指标:

1. 标准的 $M/M/1$ 模型中最优服务率为

$$\mu^*=\lambda+\sqrt{\frac{C_W}{C_S}\lambda},$$

式中 C_S 表示当 $\mu=1$ 时服务机构单位时间的费用, C_W 表示每个顾客在系统停留单位时间的费用.

2. $M/M/1(N)$模型中最优服务率为

$$\mu^*=\rho^{N+1}\frac{N-(N+1)\rho+\rho^{N+1}}{(1-\rho^{N+1})^2}=\frac{C_S}{G},$$

式中 C_S,G,λ,N 都是给定的. G 表示每服务 1 人能收入 G 元,则单位时间收入的期望值是 $\lambda(1-P_N)G$ 元,P_N 表示被拒绝的概率,$1-P_N$ 表示能接受服务的概率,$\lambda(1-P_N)$表示单位时间实际进入服务机构顾客的平均数,在稳定状态下,它也等于单位时间内实际服务完成的平均顾客数. 纯利润

$$\begin{aligned}z&=\lambda(1-P_N)G-C_S\mu\\&=\lambda\mu G\frac{\mu^N-\lambda^N}{\mu^{N+1}-\lambda^{N+1}}-C_S\mu.\end{aligned}$$

3. $M/M/1(m)$模型中最优服务率为

$$\begin{aligned}\mu^*=&\Big\{E_{m-1}\Big(\frac{m}{\rho}\Big)E_m\Big(\frac{m}{\rho}\Big)\\&+\frac{m}{\rho}\Big[E_m\Big(\frac{m}{\rho}\Big)E_{m-2}\Big(\frac{m}{\rho}\Big)\\&-E_{m-1}^2\Big(\frac{m}{\rho}\Big)\Big]\Big\}\Big/E_m^2\Big(\frac{m}{\rho}\Big)\\=&\frac{C_S\lambda}{G},\end{aligned}$$

式中 m,G,C_S 是给定的. m 表示机器台数,各台连续运转时间服从负指数分布;G 表示单位时间每台机器运转可得收入 G 元;C_S 表示服务率为 1 时的修理费用. 平均运转台数为 $m-L_s$,单位时间纯利润为

$$\begin{aligned}z&=(m-L_s)G-C_S\mu\\&=\frac{mG}{\rho}\frac{E_{m-1}\Big(\dfrac{m}{\rho}\Big)}{E_m\Big(\dfrac{m}{\rho}\Big)}-C_S\mu,\end{aligned}$$

其中

$$E_m(x)=\sum_{k=0}^{m}\frac{x^k}{k!}e^{-x}$$

称为泊松部分和. 由上式解出 μ^* 是困难的. 通常是利用泊松分布表,通过数值计算来求得,或将上式左端(对一定的 m)作为 ρ 的函数,画出图形,对于给定的 $C_S\lambda/G$,根据图形求出 μ^*/λ.

排队系统的统计推断(statistical inference of queueing system) 统计推断理论在排队系统中的应用. 与古典统计推断理论一样,分为参数估计和假设检验两大部分. 参数估计又分为点估计和区间估计. 假设检验则又分为:

1. 分布拟合度检验. 即推断相继到达间隔时间、服务时间、等待时间是服从什么类型的分布.

2. 独立性检验. 在大多数排队模型中,总是假定相继到达间隔时间是独立同分布的,诸顾客的服务时间也是独立同分布的,而且到达间隔时间与服务时间也是独立的;对具体的排队系统,应做出这些假设是否成立的检验,而大多数排队模型,这些假设只能近似地成立.

3. 统计平衡检验. 对给定的排队系统,在考察时刻是否已达到统计平衡,或者还要经过多长时间,就可认为已基本上达到统计平衡.

4. 平稳性检验. 即相继到达间隔时间与服务时间的分布类型以及参数是否随时间变化.

存　贮　论

存贮论(inventory theory)　运筹学的一个分支.是研究物资储备控制策略的一门学科.例如,在工厂生产中,需要存贮一定数量的原料,存贮过多不仅积压资金,还要支付存贮费.在商店里,若存贮商品数量不足,发生缺货现象,会减少利润;若存量过多,造成商品积压,则使经济遭受损失.银行为了进行正常的营业,需要一定的货币余额以供周转,若余额过多,则积压资金;若余额不足,则难以满足周转需要.存贮问题是人类社会在政治、经济、军事等活动领域中普遍存在的.存贮不仅可以支持日常的经营活动,而且通过它的调节,可以满足高于平均水平的需求和防止低于平均水平的供给,有时大批量采购或者利用季节性价格波动,还可以得到价格上的优惠.因此,科学地组织存贮管理,权衡存贮量的多少,以及何时补充存量,补充多少等,是一个十分重要的问题.哈里斯(Harris,F.)于1919年对商业中的存贮问题进行了研究,建立了一个简单的确定性模型,并求得了最优解.威尔逊(Wilson,R.H.)于1934年重新得出了哈里斯的公式,称为经济订购批量公式,简称EOQ公式或威尔逊公式.1945年以来,对随机或非平稳需求的库存模型进行了研究,还研究了从订货到收货之间的时间延滞的随机性以及多品种、多级的库存管理系统等.近年来,将存贮论应用于计算机管理,在存贮的信息管理系统方面提出了许多新的研究课题.

库存论(inventory theory)　即"存贮论".

存贮问题(inventory problem)　存贮论的基本问题.指包含供、存、销三个环节的存贮系统的最优化问题.即通过订货以及到货后的库存与销售,来满足顾客的要求.在这个系统中,管理者可以通过控制订货时间的间隔及订货量来调节系统的运行,使得在某种准则下系统的性能达到最优.一般地,需要知道需求供货滞后时间、缺货处理方式、费用结构、目标函数以及采取的存贮策略.管理者通过何时订货、订多少货来对系统实施控制;在采用周期盘点时,周期长度也是一个控制变量,用连续或周期性的方式盘点货物,以便掌握库存水平.

存贮系统包括:有一个存贮点,有并行可用的多个存贮点,有串联形状的存贮点和网络形状的存贮点等.存贮问题的模型有确定性模型和随机性模型两类.确定性模型中的数据皆为确定的数值,随机性模型中含有随机变量.

需求(demand)　存贮论的基本概念之一.需求即指系统的输出.在一定时间内的需求量可以通过对历史数据的统计分析,或基于对市场销货量的调查分析而获得.需求可以是确定性的,也可以是随机性的或者是非平稳的.有时需求是间断式的,有时需求又是连续均匀的.常用随机需求量的分布有泊松分布、正态分布、伽马分布和对数正态分布等.对需求量特征和规律的掌握,是制订合理的存贮计划的重要前提.

补充(replenish)　存贮论的基本概念之一.补充是指系统的输入.补充的办法是通过订货或安排生产来获得.从订货到货物进入"存贮"的这段时间,称为滞后时间.为了在某一时刻能补充存贮,需提前订货,这段时间称为提前时间.滞后时间可能很长,也可能很短,可能是确定性的,也可能是随机性的.

存贮系统(inventory system)　见"存贮问题".

存贮策略(inventory policy)　存贮论的基本概念之一.指多少时间补充一次以及每次补充数量的策略.最常见的形式有:

1. t 循环策略.即每隔 t 时间补充存贮量 Q,有时称为定期订货方式,一般采用最佳批量公式.

2. (s,Q)策略和(s,S)策略.它们都是随时要检查存贮量.当存贮量 $x>s$ 时不补充;当 $x\leqslant s$ 时订货,补充存贮.(s,Q)型是每次订货量为 Q,Q 由最佳批量公式确定.(s,S)型是每次订货量均变化,但要求订完货后使最后存贮水平达到 $S(S=Q+x)$.

3. (T,s,S)混合策略.以周期 T 盘点,即定期检查存贮量 x.当 $x>s$ 时不补充,当 $x\leqslant s$ 时补充,使订货后存贮水平达到 S.有些国控物资,不随时订货,而规定每年或每季召开有规定的订货会.

费用(cost)　存贮论的基本概念之一.指存贮、订货、生产、缺货和控制系统等所需的支出.存贮费用是指使用仓库、保管货物、保险、税金以及货物损坏变质等支出的费用.订货费用是指订购费用(固定费用)和货物的成本费用(可变费用).订购费用包括手续费、电信往来和外出采购等费用,它与订货次数有关,而与订货数量无关.货物的成本费用包括货物本身的价格、运费等,它与订货数量有关.设货物单价(或单位成本)为 k 元,订购费用为 c_3 元,订货数量为 Q,则订货费用为

$$c(Q)=\begin{cases}c_3+kQ & (Q>0),\\ 0 & (Q=0).\end{cases}$$

生产费用是指补充存贮时,不需向外厂订货,由本厂自行生产所支出的费用.这种费用共有两项:一项是装配费用,即固定费用,或称准备、结束费用;另一项是可变费用,即与生产产品的数量有关的费用.缺货费用是当存贮供不应求时所引起的损失,诸如失去销售机会、停工待料和不履行合同而缴纳罚款等损失.在不允许缺货时,以无穷大表示缺货费用.存贮论中常用的目标函数有平均费用(或利润),以及折扣费用(或利润).相应的最优策略使费用达到极小

或利润达到极大.

确定性存贮模型Ⅰ(deterministic inventory model Ⅰ) 一种存贮模型.指不允许缺货,生产时间很短的一类存贮模型.其基本假定为:

1. 单品种连续盘点.
2. 不允许缺货,即缺货费用为无穷大.
3. 当存贮降至零时,可以立即得到补充(即生产时间或滞后时间近似为0).
4. 需求是连续的、均匀的,需求率(单位时间的需求量)R 为常数,t 时间(或周期)的需求量为 Rt.
5. 采用(s,S)策略.
6. 每次订货量不变,订购费不变(每次生产量不变,装配费不变),单位存贮费不变.

设订货量为 $Q=Rt$,订货费为 c_3,货物单价为 k,则订货费为 c_3+kRt. t 时间(或周期)内的平均存贮量为

$$\frac{1}{t}\int_0^t R t \mathrm{d}t=\frac{1}{2}Rt.$$

设 c_1 为单位存贮费用,则目标函数为长期运行下 t 时间内总的平均费用,即

$$c(t)=\frac{c_3}{t}+kR+\frac{1}{2}c_1Rt$$

或

$$c(t)=\frac{c_3}{t}+\frac{1}{2}c_1Rt,$$

称 $c(t)$为费用函数.在这些基本假定下,问题为求 t 使 $c(t)$达到最小.显然每隔

$$t_0=\sqrt{\frac{2c_3}{c_1R}}$$

时间(或周期)订货一次,可使 $c(t)$最小,这里的 t_0 是最优解.订货批量

$$Q_0=Rt_0=R\sqrt{\frac{2c_3}{c_1R}}=\sqrt{\frac{2c_3R}{c_1}}.$$

费用函数(cost function) 一种特殊函数.指确定性存贮模型Ⅰ中的目标函数,即

$$c(t)=\frac{c_3}{t}+kR+\frac{1}{2}c_1Rt$$

或

$$c(t)=\frac{c_3}{t}+\frac{1}{2}c_1Rt$$

(参见“确定性存贮模型Ⅰ”).最佳费用是

$$c(t_0)=\min c(t)=\sqrt{2c_1c_3R}.$$

经济订购批量公式(economic ordering quantity formula) 存贮论的重要公式.是为使总平均费用达到最小,在每个订货期订货的数量公式.关于确定性存贮模型Ⅰ,经济订购批量为

$$Q_0=Rt_0=\sqrt{\frac{2c_3R}{c_1}}.$$

上述式子简称EOQ公式,或经济批量公式.EOQ公式是存贮论中的基本模块.

确定性存贮模型Ⅱ(deterministic inventory model Ⅱ) 一种存贮模型.是不允许缺货,生产需一定时间的一类存贮模型.其基本假定除生产时间(或周期)$T\neq0$ 外,其余的假定都与确定性存贮模型Ⅰ相同.因此,生产率

$$P=\frac{Q}{T},\quad T=\frac{Q}{P}=\frac{Rt}{P};$$

t 时间内的平均存贮量为$(P-R)Tt/2$;t 时间内所需存贮费为 $c_1(P-R)Tt/2$;t 时间内所需装配费为 c_3;目标函数为长期运行下 t 时间内平均总费用,即

$$c(t)=\frac{1}{t}\left[\frac{1}{2}c_1(P-R)Tt+c_3\right]=\frac{1}{t}\left[\frac{1}{2}c_1(P-R)\frac{Rt^2}{P}+c_3\right];$$

最佳周期为

$$t_0=\sqrt{\frac{2c_3P}{c_1R(P-R)}};$$

订货批量为

$$Q_0=\sqrt{\frac{2c_3RP}{c_1(P-R)}};$$

最佳费用为

$$c(t_0)=\min c(t)=\sqrt{2c_1c_3R\frac{(P-R)}{P}};$$

最佳生产时间为

$$T_0=\frac{Rt_0}{P}=\sqrt{\frac{2c_3R}{c_1P(P-R)}}.$$

确定性存贮模型Ⅲ(deterministic inventory model Ⅲ) 一种存贮模型.是允许缺货、生产时间很短的一类存贮模型.其基本假定除允许缺货外,其余的都与确定性存贮模型Ⅰ相同.设最初存贮量为 S,可以满足 t_1 时间的需求,即 $S=Rt_1$. t_1 时间的平均存贮量为 $S/2$,在$(t-t_1)$时间的存贮量为零;平均缺货量为 $R(t-t_1)/2$;在 t_1 时间内所需存贮费为

$$c_1\frac{1}{2}St_1=\frac{1}{2}c_1\frac{S^2}{R};$$

在 t 时间内的缺货费为

$$c_2\frac{1}{2}R(t-t_1)^2=\frac{1}{2}c_2\frac{(Rt-S)^2}{R}$$

(c_2 为缺货费);

目标函数为平均总费用,即

$$c(t,S)=\frac{1}{t}\left[c_1\frac{S^2}{2R}+c_2\frac{(Rt-S)^2}{2R}+c_3\right]$$

(c_3 为订购费);

最佳周期为

$$t_0=\sqrt{\frac{2c_3(c_1+c_2)}{c_1Rc_2}};$$

最初最佳存贮量为

$$S_0=\sqrt{\frac{2c_2c_3R}{c_1(c_1+c_2)}};$$

订货量为

$$Q_0=Rt_0=\sqrt{\frac{2Rc_3}{c_1}\frac{c_1+c_2}{c_2}};$$

在 t_0 时间内的最大缺货量为

$$Q_0-S_0=\sqrt{\frac{2Rc_3c_1}{c_2(c_1+c_2)}};$$

最佳费用为

$$c(t_0,S_0)=\min c(t,S)=\sqrt{\frac{2c_1c_2c_3R}{c_1+c_2}}.$$

确定性存贮模型Ⅳ(deterministic inventory model Ⅳ) 一种存贮模型.是需求非平稳的有限阶段的一类存贮模型.其基本假定除单品种周期盘点、存贮系统在 T 个有限的周期中运行,需求$\{d(t),t=\overline{1,T}\}$可随 t 变化且为已知,货物在一个周期刚开始时得到补充,在第 t 个周期开始时补充量为 $p(t)(\geqslant 0)$,$t=\overline{1,T}$外,其余与确定性存贮模型Ⅰ相同.目标函数为确定一串非负采购量$\{p(t),t=\overline{1,T}\}$,使有限时段 T 中的总费用最小.这类模型可用最优化算法求解.这种算法还可以推广到滞后时间是常数或非负整数随机变量、每个周期中的需求是非负随机变量、缺货可以延期补足等模型.由于此法计算复杂,一般采用启发式算法.

确定性存贮模型Ⅴ(deterministic inventory model Ⅴ) 一种存贮模型.即多品种多级存贮模型.在此系统中,最高一级(第一级)是货源点,它供应下一级(第二级)的库存点.依次下去,最低级的库存点(如零售点)直接满足顾客的需求.订货策略可能各异:单个物品分别订货;联合订货,即同时订购所有品种;混合式订货,它允许订购所有或部分物品.相应于这三种订货策略,订购费用对最优策略有很大影响.对多品种多级存贮问题的分析,近年来有所进展.

报童问题(newsboy problem) 一个需求是随机的典型存贮问题.问题是:一个报童每售出一份报纸赚 k 元,未售出每份赔 h 元,每日售出 r 份报纸的概率$P(r)$是已知的,问报童每日应准备多少份报纸为最好?即报童每日报纸的订货量 Q 为何值时,赚钱的期望值最大?换言之,即如何适当地选择 Q 值,使因不能出售而受到的损失以及因缺货失去销售机会的损失、两者期望值之和最小.当 $r<Q$(供过于求),此时因报纸不能全部售出而承担损失,其期望值为

$$\sum_{r=0}^{Q}h(Q-r)P(r);$$

当 $r>Q$(供不应求),此时因缺货少赚钱而损失,其期望值为

$$\sum_{r=Q+1}^{\infty}k(r-Q)P(r).$$

因此,损失的期望值为

$$c(Q)=h\sum_{r=0}^{Q}(Q-r)P(r)+k\sum_{r=Q+1}^{\infty}(r-Q)P(r).$$

最优的 Q 应满足 $c(Q)\leqslant c(Q+1)$和 $c(Q)\leqslant c(Q-1)$,由此可从下式求出 Q:

$$\sum_{r=0}^{Q-1}P(r)\leqslant\frac{k}{k+h}\leqslant\sum_{r=0}^{Q}P(r).$$

随机存贮模型Ⅰ(stochastic inventory model Ⅰ) 一种存贮模型.指需求是随机离散的存贮模型.此模型只解决一次订货问题(因为约定两次订货之间没有联系,都作为独立的一次订货).称此单品种单周期的随机存贮模型为报童问题(参见“报童问题”).此模型可推广到批量订货有折扣的情形.一般地,设 Q 为订货量,订购费为 c_3,y 为决策后的库存水平,$L(y)$是库存水平为 y 的条件下存贮与缺货的期望总费用,则有

$$L(y)=c_2(d-y)+(c_2+c_1)\int_0^y\varphi(t)\mathrm{d}(t),$$

式中 c_1 和 c_2 分别为周期结束时每个现货的存贮费和缺货费,$\varphi(t)$为一个周期中的需求量 r 的分布$(0\leqslant t<+\infty)$,平均需求 $d<+\infty$.若令 x 为决策前的初始存贮量,$c(x)$为采用最优决策后一个周期的最小期望费用,于是

$$c(x)=\min_{y\geqslant x}\{c_3(y-x)+L(y)\}.$$

记$G(y)=cy+L(y)$,则最优策略是(s,S)型的,其中 S 的最优值 S^* 使 $G(x)$达到最小,$G(x)$为凸函数,S^*惟一,而 S 的最优值 S^* 是方程 $k+G(S^*)=G(S^*)$的解.

随机存贮模型Ⅱ(stochastic inventory model Ⅱ) 一种存贮模型.指需求是随机离散的(s,S)型存贮模型.假定:

1. 原有存贮量为 I(在本阶段内为常数).
2. 需求量为 r,其概率 $P(r)$为已知,$\sum P(r)=1$.

当本阶段开始时订货量为 Q,存贮量达到 $S=I+Q$ 表示存贮所达到的水平,订货费为 c_3+kQ.对于存贮费,当需求 $r<S$ 时,未能售出的存贮部分需付存贮费;当需求 $r\geqslant S$ 时,不需要付存贮费.所需存贮费的期望值为

$$\sum_{r\leqslant S}c_1(S-r)P(r)$$

$(r=S$ 时,不付存贮费及缺货费$)$.当需求 $r>S$ 时,$(r-S)$部分需付缺货费.缺货费用的期望值为

$$\sum_{r>S} c_2(r-S)P(r).$$

问题是使本阶段所需订货费及存贮费、缺货费期望之和为最小. 期望总费用为

$$c(S) = c_3 + k(S-I) + \sum_{r \leqslant S}(S-r)c_1P(r) + \sum_{r>S}c_2(r-S)P(r).$$

随机存贮模型Ⅲ(stochastic inventory model Ⅲ) 一种存贮模型. 指需求和滞后时间都是随机离散的存贮模型. 假定 t 时间内的需求量 r 是随机的, 若概率 $\varphi_t(r)$ 和单位时间内的平均需求量 ρ, t 时间内的平均需求量为 ρt. 设年平均需求量为 D, 缓冲(安全)存贮量为 B. 问题是如何确定缓冲存贮量 B、定货点 L 以及订货量 Q_0, 使总费用达到最小? 订货量为

$$Q_0 = \sqrt{\frac{2c_3D}{c_1}};$$

定货点 $L=D_L+B=\mu\rho+B(D_L)$ 表示滞后时间内的平均需求量(μ 表示平均滞后时间);

$$P_L = \sum_x P(x) \cdot F_x(L)$$

表示缺货的概率($P(x)$ 表示滞后时间为 x 天的概率,

$$F_x(L) = \sum_{r>L}\varphi_x(r)$$

表示订货点为 L 而在 x 天内需求 $r>L$ 的概率); 缺货费的期望值为 c_2P_L; n_0 次缺货费的期望值为 $n_0c_2P_L$; 每年的存贮费用为 $(Q_0/2+B)c_1$. 因 Q_0 为最佳值, 故只需考虑缓冲存贮量的存贮费用以及缺货费期望值两者之和最小. 即令 $n_0c_2P_L+c_1B$ 为最小, 可以确定 L 和 B. 由关系式 $L=\mu\rho+B$, 只要确定了 L, 就可以确定 B.

随机存贮模型Ⅳ(stochastic inventory model Ⅳ) 一种存贮模型. 指单品种多阶段随机存贮模型. 假定: 采用 (T,s,S) 策略(参见“存贮策略”). 滞后时间为 0, 缺货在以后补充. 令 $t=1,2,\cdots$, 需求量为 r, 周期 t 中的需求量为 $r(t)$; 周期 t 开始时订货量为 Q_0, 存贮水平为 $Q_0(t)$; 周期 t 结束时存贮水平为 $Q_n(t)$. 设 $r(t)(t=1,2,\cdots)$ 独立同分布, 平均需求 d 有限, 有概率密度 $\varphi(t)(t\geqslant 0)$. 在 (T,s,S) 策略下有

$$Q_0(t+1) = \begin{cases} S & (-\infty < Q_n(t) \leqslant s), \\ Q_0(t)-r(t) & (s<Q_n(t)<S). \end{cases}$$

因此, $\{Q_0(t), t=1,2,\cdots\}$ 是一时间离散状态连续的马氏链. 问题归为求订货费、存贮费及缺货费期望之和最小. 期望总费用为

$$c(Q,S) = c_3d + \frac{1}{\Delta(Q)}\left[c_3 + L(S) + \int_0^Q L(S-x)m(x)\mathrm{d}x\right],$$

其中 $Q=S-s$, $L(y)$ 由式

$$c_2(d-y) + (c_2+c_1)\int_0^y \varphi(t)\mathrm{d}t$$

给出, 而

$$m(t) = \sum_{n=1}^{\infty}\varphi_n(t),\ \Delta(Q) = \int_0^Q m(x)\mathrm{d}x,$$

φ_n 则为 φ 对其自身的 n 重卷积, $\varphi_1=\varphi_0$. 由 $c(Q,S)$ 可算出 Q, S 的最优值.

***ABC* 分类管理法**(*ABC* classification management method) 对存贮物资分成 *ABC* 三类分类管理的方法. *ABC* 三类的划分并不是绝对的, 一般根据物资占用存贮金额的总费用比例来划分. *A* 类物资品种少(约占 10%~20%), 而所占总存贮费用最多(约占 60%~80%); *C* 类品种最多(约占 50%~70%), 而所占总存贮费用最少(约占 5%~10%); *B* 类在 *A* 与 *C* 两类之间. 对 *A* 类物资应该使用仔细的存贮控制公式而且在管理上特别注意, 对 *B* 类可粗略一些, 对于 *C* 类一般可用经验来确定.

在生产中, 有些物料需要根据产品总计划倒推出来, 不同时期的需求量也不同, 问题在于决定如何订货. 计划工作的步骤为:

1. 计算净需求量.
2. 确定生产批量.
3. 确定订货合同下达日期.
4. 零部件展开.

产品往往可以分解为一些零部件, 而各种零部件所需数量是不一样的. 另外, 由于各种零部件产品质量不同, 组装时的损失也不同, 因此, 考虑这些因素对产品及这些零部件的净需求量自然也不相同.

易腐物品存贮理论(inventory theory for perishable material) 研究易于腐烂物品的存贮理论. 如血液、食品、药品、武器弹药等方面都是易腐物品存贮理论的重要应用领域. 易腐物品存贮模型是按寿命来分类的, 可分为固定寿命和随机寿命两类.

易腐存贮模型与排队模型间有如下关系: 易腐存贮中的存贮水平相当于排队模型中的队长, 需求相当于服务过程, 物品寿命相当于逗留时间, 存贮中新鲜物品的补充相当于顾客到达. 但两者之间也有重要的区别: 在存贮问题中, 何时订货, 订多少是控制变量; 而在排队问题中, 顾客的到达是随机的.

概率约束存贮模型(inventory model with probability constraint) 一类存贮模型. 此模型的背景描述如下: 设工厂 B 以固定的速度消耗某种原材料. 某季度开始以前, B 与 A 签订合同, 该季度 B 所需原材料由 A 提供, A 可分若干次交货, 但是具体交货时间不能事先确定. 为保证生产顺利进行, B 在季度开始时需要有一定数量的原材料储备, 那么应储备多少才能维持正常生产? 模型描述如下: 考虑

在$(0,T)$固定的区间内，假定交货次数n取定且设：

1. B厂在$(0,T)$中连续使用这些原材料，单位时间的用量c是常数.

2. 交货时刻是随机的，假定是$(0,T)$中独立同分布的n个均匀随机变量.

3. 每次的交货量相同，为cT/n.

取B的初始存贮水平M为模型的决策变量，$X(t)$表示时刻t以前的累积交货量，则有

$$P\{\inf_{0\leqslant t\leqslant T}(M+X(t)-ct)\geqslant 0\}=1-\varepsilon,$$

这里ε是事先指定的$(0,1)$中的数. 上式等价于

$$P\left\{\sup_{0\leqslant t\leqslant T}\left(\frac{t}{T}-\frac{1}{cT}X(t)\right)\leqslant\frac{M}{cT}\right\}=1-\varepsilon.$$

利用$(0,T)$中均匀随机变量顺序量的性质，可以求得左端的精确结果以及极限结果

$$\lim_{n\to\infty}P\left\{\sqrt{n}\sup_{0\leqslant t\leqslant T}\left(\frac{t}{T}-\frac{1}{cT}X(t)\right)\leqslant y\right\}$$
$$=1-\exp(-2y^2)\quad(y\geqslant 0).$$

于是，可得实用上方便的近似公式

$$M\approx cT\sqrt{\frac{1}{2n}\lg\frac{1}{\varepsilon}}.$$

假定B厂生产增加时原材料的消耗及交货次数n按比例增加，即

$$\frac{n}{c}=\text{常数},$$

那么$M\approx a\sqrt{c}$，a为某常数. 因而，M对c的弹性

$$\frac{\mathrm{d}M}{\mathrm{d}c}\Big/\left(\frac{M}{c}\right)=\frac{1}{2}.$$

由此，若B厂的生产增加1%，则其初始库存水平只需增加0.5%.

现金管理存贮模型(inventory model of cash management)　一类特殊存贮模型. 指在银行或企业中，为保障现金的交易、应急及投机等需求而建立的现金贮备优化管理模型. 用$\mu_n+\xi_n$表示第n天的日净现金流，其中ξ_n是均值为0的独立随机变量，μ_n为现金流的均值，$n=1,2,\cdots,N$，N为平均期限的天数(做出决策前的现金水平称为开结余，一天营业结束后的现金水平称为闭结余)，即对N天计算平均结余. 第n步的开结余为B_n，

$$S_n=\sum_{i=n}^{N-1}B_i$$

是从开始到第n步以前的日累积闭结余，约定$S_N=0$. 现金兑换费用具有如下的结构：

$$T(Z)=\begin{cases}K_u+c_uZ & (Z>0),\\ 0 & (Z=0),\\ K_d-c_dZ & (Z<0),\end{cases}$$

式中$c_u,c_d>0$，$K_u,K_d\geqslant 0$为常数，Z是现金兑换量，$Z>0$表示变卖资产使现金增加，$Z<0$表示购入资产以减少现金. 由式

$$y_n(x)=\begin{cases}U_n & (x\leqslant u_n),\\ x & (u_n<x<d_n),\\ D_n & (d_n\leqslant x)\end{cases}$$

确定的策略称为简单策略，其中$y_n(x_n)$是第n步做出决策之后的现金水平，$u_n\leqslant U_n\leqslant D_n\leqslant d_n$是策略的参数. 在相当广的范围内，该策略就是最优策略. 现金管理中的存贮模型与一般存贮模型有两点不同：

1. 在一个周期中现金的库存水平不但有可能下降，而且有可能上升，这将由收入小于或大于支出所决定；同样地，在周期开始时，决策者改变库存水平也可以有两个方向：减少现金或增加现金.

2. 在一般的库存模型中，不管是缺货还是过剩，都是以0为基准的. 而对现金管理问题，则有所谓补偿结余的要求. 一是绝对补偿结余要求，要求每个周期都有一定的结余；二是平均补偿结余要求，要求在包括若干周期的某个时期中，按周期平均的结余等于某个给定值，且平均补偿结余要求是合理的.

对 策 论

对策论(game theory)　亦称博弈论. 运筹学的一个重要分支. 是研究在相互具有竞争和对抗的体系中，各方如何选取策略，以使最后能得到对己方最有利结果的一门学科. 对策论是在研究赌博问题的基础上发展起来的，后来不断扩展到军事、社会和经济领域，逐渐形成一门研究互动关系的数学理论. 在中国，历史上著名的“齐王和田忌赛马”是对策的一个典型例子. 1921年，德国数学家策梅洛(Zermelo, E. F. F.)用集合论的方法研究了国际象棋的着法，发表了“关于集合论在象棋对策中的应用”一文，首次把对策问题置于严密的数学理论与方法的基础之上. 同年，法国数学家波莱尔(Borel, (F.-É.-J.-)É.)讨论了对策论中的一些个别现象，引入了最优策略的概念. 1928年，冯·诺伊曼(von Neumann, J.)证明了两人零和对策的主要结果. 二次世界大战期间，由于军事上的需要，使“对策现象”开始成为应用数学的重要研究对象. 1944年，冯·诺伊曼和莫根施特恩(Morgenstern, O.)把这些方面的数学理论加以系统和完善，出版了《对策论和经济行为》一书，使对策论的研究开始系统化和公理化. 20世纪50年代，纳什(Nash, J. F.)和夏普利(Shapley, L. S.)等在非合作对策和合作对策方面做出了重要贡献，使对策论的理论和应用研究发展到一个新的阶段. 对策论在管理科学、数量经济学、统计决策，以及军事计划中都有重要应用. 对策论的许多术语和概念，则已用于更广泛的科学和技术领域.

对策论分为静态对策和动态对策两大类. 静态对策分结盟对策和不结盟对策两种. 结盟对策又分

为联合对策和合作对策两种;不结盟对策则分为有限对策和无限对策.无论是有限还是无限对策,皆有两人对策和多人对策;无论是两人对策还是多人对策,皆有零和对策和非零和对策之分.

博弈论(game theory) 即"对策论".

对策现象(game phenomenon) 对策论的概念之一.指对策问题中带有竞争性质(或至少含有竞争成分)的现象.

齐王与田忌赛马(horse race between the Qi king and Tianji) 一个安排马匹比赛的典型对策问题.指中国战国时期,齐王与田忌赛马的故事:各人备上、中、下三个等级的马各一匹,每匹都得参加比赛且只参加一次,每次比赛各出一匹马,一共比赛三次,每次比赛后负者要付给胜者一千金.当时齐王的马按等级依次略强于田忌的马,但各劣于田忌上一等级的马.田忌的谋士孙膑给田忌出了个主意:每次比赛先让齐王说出他要出哪一匹马,并让田忌用下马对齐王的上马,用中马对齐王的下马,用上马对齐王的中马.比赛结果,田忌二胜一负而得一千金.

对策问题三要素(three key elements of game problem) 对策论的三个基本概念.指对策问题中的局中人、策略和对策的得失.

局中人是指在一局对策(竞争或斗争)中,有决策权的参加者.局中人可以是个人,也可是集团.局中人一般用记号Ⅰ,Ⅱ,…表示,而用字母I表示所有局中人构成的集合.只有两个局中人的对策,称为两人对策.多于两个局中人的对策,称为多人对策.策略是局中人在对策中采用的一个可行的自始至终通盘筹划的行动方案.局中人策略的全体称为策略集合,它可以是有限集,也可以是无限集.局中人在一局对策结束时的得失称为对策的得失.一局对策结束时,每个局中人的"得失"是全体局中人所取定的一组策略的函数,称为支付函数,亦称赢得函数.称赢得函数为连续函数的对策为连续对策.从每个局中人的策略集中各取一个策略组成的策略组,称为对策的一个局势.对策的得失是局势的函数.

局中人(player) 见"对策问题三要素".

两人对策(two-person game) 见"对策问题三要素".

多人对策(n-person game) 见"对策问题三要素".

策略(strategy) 见"对策问题三要素".

策略集合(strategy set) 见"对策问题三要素".

对策得失(profit and loss of game) 见"对策问题三要素".

支付函数(payoff function) 见"对策问题三要素".

赢得函数(win function) 见"对策问题三要素".

局势(situation) 见"对策问题三要素".

连续对策(continuous game) 见"对策问题三要素".

两人有限零和对策(two-person finite zero-sum game) 一类对策.指在一局对策中,只有两个局中人,他们的策略都是有限的,他们的得失之和总为零.记为$\{S_1,S_2;A\}$,其中S_1和S_2分别是两个局中人的策略集合,A是支付矩阵(参见"矩阵对策").

矩阵对策(matrix game) 亦称有限对抗对策.即"两人有限零和对策".矩阵对策是对策论中最重要和最基本的对策.在矩阵对策中,将两个局中人在每个局势下的得失用矩阵形式表示出来,称此矩阵为支付矩阵,亦称赢得矩阵.矩阵对策也因此而得名.在双方进行对策时所依据的对策原则是:立足于最坏情况,去争取最好的结果.但在实际中,特别在军事对策中,智勇双全的军事家在知己知彼的情况下,攻其(对方)不备,以奇取胜的事例也是常有的.

有限对抗对策(finite confrontation game) 即"矩阵对策".

矩阵对策的数学模型(mathematical model of matrix game) 矩阵对策的数学表达形式.指对策模型

$$G=\{S_1,S_2;A\},$$

式中$A=(a_{ij})_{m\times n}$称为支付矩阵(参见"矩阵对策"),$S_1=\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$称局中人Ⅰ的纯策略集合($\alpha_1,\alpha_2,\cdots,\alpha_m$都是局中人Ⅰ的纯策略),$S_2=\{\beta_1,\beta_2,\cdots,\beta_n\}$称局中人Ⅱ的纯策略集合($\beta_1,\beta_2,\cdots,\beta_n$都是局中人Ⅱ的纯策略).局中人Ⅰ从$S_1$中选一个$\alpha_i(i=1,2,\cdots,m)$,同时局中人Ⅱ从$S_2$中选一个$\beta_j(j=1,2,\cdots,n)$,构成对策的一个局势$(\alpha_i,\beta_j)$,称$(\alpha_i,\beta_j)$为纯局势.共有$m\times n$个纯局势对应于一个支付矩阵$A=(a_{ij})_{m\times n}$.在对策$G$中,若$S_i(i\in I)$和$A$都依赖于某个随时间变化的量,称$G$为动态对策,即$G$是其策略集和赢得矩阵随时间变化的对策;否则,称$G$为静态对策.

纯策略(pure strategy) 见"矩阵对策的数学模型".

非随机策略(non-random strategy) 见"矩阵对策的数学模型".

纯策略集合(pure strategy set) 见"矩阵对策的数学模型".

纯局势(pure situation) 见"矩阵对策的数学模型".

动态对策(dynamic game) 见"矩阵对策的数学模型".

静态对策(static game)　见“矩阵对策的数学模型”.

最优策略(optimal strategy)　亦称最优纯策略. 局中人的策略之一. 是能使局中人在对策中“立足于最坏情况,争取最好结果”的那个策略. 在矩阵对策 $G=\{S_1,S_2;A\}$ 中,设 $a_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 为支付矩阵的元素,则称 $\max\limits_i\min\limits_j a_{ij}$ 对应的局中人Ⅰ的策略 α_i^* 为局中人Ⅰ的最优策略,而称 $\min\limits_j\max\limits_i a_{ij}$ 对应的局中人Ⅱ的策略 β_j^* 为局中人Ⅱ的最优策略. 其中,$\max\limits_i\min\limits_j a_{ij}$ 的意义是:在支付矩阵 A 的每一行中取最小值 $\min\limits_j a_{ij}(i=1,2,\cdots,m)$,再从这些最小值中取最大值 $\max\limits_i\min\limits_j a_{ij}$. $\min\limits_j\max\limits_i a_{ij}$ 表示:先在 A 中的每一列取最大值 $\max\limits_i a_{ij}(j=1,2,\cdots,n)$,再从这些最大值中取最小值 $\min\limits_j\max\limits_i a_{ij}$.

当局中人有最优策略 α_i^* 和 β_j^* 时,一定有

$$\max_i\min_j a_{ij}=\min_j\max_i a_{ij}=V,$$

则称 V 为矩阵对策 $G=\{S_1,S_2;A\}$ 的最优值或对策值,称该二最优策略构成的策略对 (α_i^*,β_j^*) 为对策 G 的最优局势. 并称 (α_i^*,β_j^*) 为对策 G 的鞍点,或对策 G 的纯策略解,简称对策的解. 也即对策的解是

$$\min_j\max_i a_{ij}=\min_j a_{i^*j}=\max_i a_{ij^*}=\max_i\min_j a_{ij}=V$$

的纯局势 (α_i^*,β_j^*). 矩阵对策有解的充分必要条件是:存在一个纯局势 (α_i^*,β_j^*),使对一切 $i=1,2,\cdots,m$ 和一切 $j=1,2,\cdots,n$,皆有 $a_{ij^*}\leqslant a_{i^*j^*}\leqslant a_{i^*j}$.

求解一个矩阵对策 $G=\{S_1,S_2;A\}$,只需按照最小最大原则 $(\min\limits_j\max\limits_i a_{ij})$ 和最大最小原则 $(\max\limits_i\min\limits_j a_{ij})$,分别求出局中人Ⅰ和Ⅱ的最优纯策略 α_i^* 和 β_j^*,就可立即求出对策 G 的纯策略解 (α_i^*,β_j^*) 和最优值(即对策值)$V=\min\limits_j\max\limits_i a_{ij}=\max\limits_i\min\limits_j a_{ij}$.

最优纯策略(optimal pure strategy)　见“最优策略”.

最优局势(optimal situation)　见“最优策略”.

对策值(game value)　见“最优策略”.

对策鞍点(saddle point of game)　见“最优策略”.

混合策略(mixed strategy)　亦称随机策略,简称策略. 纯策略的推广. 设纯策略集合

$$S_1=\{\alpha_1,\alpha_2,\cdots,\alpha_m\}\text{和}S_2=\{\beta_1,\beta_2,\cdots,\beta_n\}$$

对应的概率向量分别为

$$x=(x_1,x_2,\cdots,x_m)$$

$$\left(x_i\geqslant 0,i=1,2,\cdots,m;\sum_{i=1}^m x_i=1\right)$$

和

$$y=(y_1,y_2,\cdots,y_n)$$

$$\left(y_j\geqslant 0,j=1,2,\cdots,n;\sum_{j=1}^n y_j=1\right),$$

称它们分别为局中人Ⅰ和Ⅱ的混合策略,其中 x_i 表示局中人Ⅰ选用策略 $\alpha_i\in S_1$ 的概率,y_j 表示局中人Ⅱ选用策略 $\beta_j\in S_2$ 的概率. 混合策略 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的含义是:在一局对策中,对策的双方不是自始至终只分别取其同一个纯策略,而是局中人Ⅰ分别以概率 $x_1,x_2,\cdots,x_m$ 取其纯策略 $\alpha_1,\alpha_2,\cdots,\alpha_m$;局中人Ⅱ分别以概率 $y_1,y_2,\cdots,y_n$ 取其纯策略 $\beta_1,\beta_2,\cdots,\beta_n$. 对于矩阵对策 $G=\{S_1,S_2;A\}$,$S_1=\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$,$S_2=\{\beta_1,\beta_2,\cdots,\beta_n\}$,$A=(a_{ij})_{m\times n}$ 都是给定的. 纯策略是混合策略的特殊情况,即

$$x_i=\begin{cases}1 & (i=k),\\ 0 & (i\neq k),\end{cases}\qquad y_j=\begin{cases}1 & (j=s),\\ 0 & (j\neq s).\end{cases}$$

也可写为

$$\boldsymbol{x}=(0,0,\cdots,x_k,0,\cdots,0)=(0,0,\cdots,1,0,\cdots,0),$$
$$\boldsymbol{y}=(0,0,\cdots,y_s,0,\cdots,0)=(0,0,\cdots,1,0,\cdots,0).$$

将每个局中人各取其混合策略之一时构成的混合策略组 $(x_i,y_j)(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 称为混合局势(参见“对策问题三要素”).

随机策略(random strategy)　即“混合策略”.

混合局势(mixed situation)　见“混合策略”.

混合策略单纯形(simplex of mixed strategy)　混合策略集的几何表示. 由概率向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 分别构成的集合 S_1 和 S_2. 即:概率向量

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_m)$$

$$\left(x_i\geqslant 0,i=1,2,\cdots,m;\sum_{i=1}^m x_i=1\right)$$

和

$$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)$$

$$\left(y_j\geqslant 0,j=1,2,\cdots,n;\sum_{j=1}^n y_j=1\right)$$

分别构成的集合 $S_1=\{\boldsymbol{x}\}$ 和 $S_2=\{\boldsymbol{y}\}$,称它们为 $(m-1)$ 维和 $(n-1)$ 维单纯形. 单纯形 S_1 可由基向量

$$\underbrace{(1,0,0,\cdots,0)}_{m\text{个分量}},\underbrace{(0,1,0,\cdots,0)}_{m\text{个分量}},\cdots,\underbrace{(0,0,0,\cdots,1)}_{m\text{个分量}}$$

的凸组合构成;单纯形 S_2 可由基向量

$$\underbrace{(1,0,0,\cdots,0)}_{n\text{个分量}},\underbrace{(0,1,0,\cdots,0)}_{n\text{个分量}},\cdots,\underbrace{(0,0,0,\cdots,1)}_{n\text{个分量}}$$

的凸组合构成.

混合扩充(mixed expansion)　矩阵对策模型的推广. 对于矩阵对策 $G=\{S_1,S_2;A\}$,称

$$E(x,y)=\sum_{i=1}^m\sum_{j=1}^n a_{ij}x_iy_j$$

为局中人Ⅰ赢得的数学期望,简称赢得期望;称

$$-E(x,y)=-\sum_{i=1}^m\sum_{j=1}^n a_{ij}x_iy_j$$

为局中人Ⅱ的赢得期望.

设 S_1^* 和 S_2^* 分别是局中人Ⅰ和局中人Ⅱ的所有混合策略构成的集合，E 为局中人Ⅰ的赢得期望，则称 $G^*=\{S_1^*,S_2^*;E\}$ 为矩阵对策 G 的混合扩充.

混合扩充中对策的值(value of game in the mixed expansion)　对矩阵对策 G，设 S_1^* 和 S_2^* 分别是局中人Ⅰ和Ⅱ的所有混合策略的集合，称

$$V=\min_{y\in S_2^*}E(x^*,y)=\min_{y\in S_2^*}\max_{x\in S_1^*}E(x,y)$$
$$=\max_{x\in S_1^*}\min_{y\in S_2^*}E(x,y)=\max_{x\in S_1^*}E(x,y^*)$$

为对策 G 的值，x^* 和 y^* 分别为局中人Ⅰ和Ⅱ的最优混合策略，(x^*,y^*) 为 G 在混合策略(意义)下的解. 对于混合扩充，有以下基本定理：混合扩充中任何一个给定的矩阵对策 G 一定有解. 即任一个矩阵对策一定存在混合策略(意义)下的解.

混合策略下的解(solution under the mixed strategy)　见“混合扩充中对策的值”.

对策的基本定理(fundamental theory of game)　见“混合扩充中对策的值”.

混合扩充中的最优策略(optimal strategy in the mixed expansion)　见“混合扩充中对策的值”.

矩阵对策的线性规划解法(solving method of linear programming of matrix game)　把矩阵对策问题化为求解线性规划问题的方法. 把解不等式组

$$\begin{cases}\sum_{i=1}^{m}a_{ij}x_i\geqslant V & (j=1,2,\cdots,n),\\ \sum_{i=1}^{m}x_i=1, & \\ x_i\geqslant 0 & (i=1,2,\cdots,m)\\ \quad (V=\max_{x\in S_1^*}\min_{1\leqslant j\leqslant n}\sum_{i=1}^{m}a_{ij}x_i) & \end{cases}$$

化为求解极小化形式的线性规划问题：

$$\begin{cases}\min\ \sum_{i=1}^{m}x'_i, & \\ \text{s.t.}\ \sum_{i=1}^{m}a_{ij}x'_i\geqslant 1 & (j=1,2,\cdots,n),\\ \quad x'_i\geqslant 0 & (i=1,2,\cdots,m),\end{cases}$$

其中

$$x'_i=\frac{x_i}{V}(i=1,2,\cdots,m).$$

同样，把求解不等式组

$$\begin{cases}\sum_{j=1}^{n}a_{ij}y_j\leqslant V & (i=1,2,\cdots,m),\\ \sum_{j=1}^{n}y_j=1, & \\ y_j\geqslant 0 & (j=1,2,\cdots,n)\\ \quad (V=\min_{y\in S_2^*}\max_{1\leqslant i\leqslant m}\sum_{j=1}^{n}a_{ij}y_j) & \end{cases}$$

化为求解极大化形式的线性规划问题：

$$\begin{cases}\max\ \sum_{j=1}^{n}y'_j, & \\ \text{s.t.}\ \sum_{j=1}^{n}a_{ij}y'_j\leqslant 1 & (i=1,2,\cdots,m),\\ \quad y'_j\geqslant 0 & (j=1,2,\cdots,n),\end{cases}$$

其中 $y'_j=\frac{y_j}{V}(j=1,2,\cdots,n)$.

矩阵对策的解法(solving method of matrix game)　求解矩阵对策的方法. 这里特别强调求解矩阵对策在混合策略意义下的解与值的方法(对于矩阵对策的纯策略解，参见“最优策略”). 有以下四种基本的解法：

1. 设对策 G 没有鞍点，且 $x_i^*\neq 0$，$y_j^*\neq 0$. 把不等式组

$$\begin{cases}\sum_{i=1}^{m}a_{ij}x_i\geqslant V & (j=1,2,\cdots,n),\\ x_i\geqslant 0 & (i=1,2,\cdots,m),\\ \sum_{i=1}^{m}x_i=1 & \end{cases}$$

和不等式组

$$\begin{cases}\sum_{j=1}^{n}a_{ij}y_j\leqslant V & (i=1,2,\cdots,m),\\ y_j\geqslant 0 & (j=1,2,\cdots,n),\\ \sum_{j=1}^{n}y_j=1 & \end{cases}$$

都化成线性方程组，再求解.

2. 设对策 G 有鞍点，且 $x_i^*\neq 0$，$y_j^*\neq 0$. 根据矩阵对策的解法1，

若 $x_i^*\neq 0$，则 $\sum_{j=1}^{n}a_{ij}y_j^*=V$；

若 $y_j^*\neq 0$，则 $\sum_{i=1}^{m}a_{ij}x_i^*=V$；

若 $\sum_{j=1}^{n}a_{ij}y_j^*<V$，则 $x_i^*=0$；

若 $\sum_{i=1}^{m}a_{ij}x_i^*>V$，则 $y_j^*=0$

(这里 $V=E(x^*,y^*)$，(x^*,y^*) 是对策 G 的最优混合局势). 此时，可以把不等式变成等式来求解. 例如，给定一个对策 $G=\{S_1,S_2;A\}$，若

$$A=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix},$$

则

$$\begin{cases}a_{11}x_1+a_{21}x_2=V,\\ a_{12}x_1+a_{22}x_2=V,\\ x_1+x_2=1\end{cases}\quad 和\quad \begin{cases}a_{11}y_1+a_{12}y_2=V,\\ a_{21}y_1+a_{22}y_2=V,\\ y_1+y_2=1.\end{cases}$$

分别解这两个方程组，即得.

3. 对于某些特殊结构的矩阵，用降低阶数的方法来求解. 设 $G=\{S_1,S_2;A\}$ 是一个矩阵对策，其中 $S_1=\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$，$S_2=\{\beta_1,\beta_2,\cdots,\beta_n\}$，$A=(a_{ij})_{m\times n}$. 若 α_1 为其余的纯策略 $\alpha_i(1<i\leqslant m)$ 之一所优超(若对于一切 $j(1\leqslant j\leqslant n)$，均有 $a_{ij}\geqslant a_{kj}$，则称局中人 Ⅰ 的纯策略 α_i 优超于 α_k. 同样，对于一切的 $i(1\leqslant i\leqslant m)$，均有 $a_{ij}\leqslant a_{il}$，则称局中人 Ⅱ 的纯策略 β_j 优超于 β_l)，由 G 可得到一个新的对策 $G'=\{S'_1,S_2;A'\}$，其中 $S'_1=\{\alpha_2,\alpha_3,\cdots,\alpha_m\}$，$a'_{ij}=a_{ij}(i=2,\cdots,m;j=1,2,\cdots,n)$，则：

1) $V_{G'}=V_G$.

2) G' 中局中人 Ⅱ 的最优策略便是 G 中局中人 Ⅱ 的最优策略.

3) 若 $(x_2^*,x_3^*,\cdots,x_m^*)$ 是 G' 中局中人 Ⅰ 的最优策略，则 $(0,x_2^*,\cdots,x_m^*)$ 是 G 中局中人 Ⅰ 的最优策略.

4. 对于某些特殊结构的矩阵，可以使其元素尽可能多地变成零. 给定两个矩阵对策

$$G_1=\{S_1,S_2;A_1=(a_{ij})\},$$
$$G_2=\{S_1,S_2;A_2=(a_{ij}+d)\},$$

其中 d 是常数，则两个对策的解集合不变，其对策值相差一个 d，即 $V_2=V_1+d$(V_2 和 V_1 分别为 G_2 和 G_1 的对策值).

对策的图解法(diagram method of game)　借助图示求解对策问题的方法. 图解法是根据 $G=\{S_1,S_2;A\}$ 中的矩阵 A，求出函数，做出相应的直线，从图象中求出使最小值函数达到最大，或求出使最大值函数达到最小. 例如，设局中人 Ⅰ 有两个纯策略 x_1 和 $x_2=1-x_1$，而局中人 Ⅱ 有 n 个纯策略 $y_1,y_2,\cdots,y_n$，

$$A=\begin{pmatrix} a_{11},a_{12},\cdots,a_{1n} \\ a_{21},a_{22},\cdots,a_{2n}\end{pmatrix}.$$

把 n 个式子

$$a_{11}x_1+a_{21}x_2=a_{21}+(a_{11}-a_{21})x_1,$$
$$a_{12}x_1+a_{22}x_2=a_{22}+(a_{12}-a_{22})x_1,$$
$$\cdots\cdots$$
$$a_{1n}x_1+a_{2n}x_2=a_{2n}+(a_{1n}-a_{2n})x_1$$

看成以 $x_1(0\leqslant x_1\leqslant 1)$ 为变量的 n 条直线. 然后，取所有这些直线对应的最低点组成的下包络(如图). 最后，相应于下包络所有点中取最大值的那个点 $\bar{x}$，即为局中人 Ⅰ 最优混合策略 x_1 的应取概率值，而 $1-\bar{x}$ 即为 x_2 应取的值.

两人无限零和对策(two-person infinite zero-sum game)　一类零和对策. 指这样的矩阵对策 $G=\{S_1,S_2;A\}$：策略集 S_1 和 S_2 中至少有一个是无限集合，得失之和总为零.

对两人无限零和对策，记

$$\begin{aligned}\overline{A}(\boldsymbol{x},\boldsymbol{y})&=E^xE^yA(\boldsymbol{x},\boldsymbol{y})=E^yE^xA(\boldsymbol{x},\boldsymbol{y})\\ &=\int_{\mathscr{X}}\int_{\mathscr{Y}}A(\boldsymbol{x},\boldsymbol{y})g_2(\boldsymbol{y})g_1(\boldsymbol{x})\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{x}\\ &=\int_{\mathscr{Y}}\int_{\mathscr{X}}A(\boldsymbol{x},\boldsymbol{y})g_1(\boldsymbol{x})g_2(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y},\end{aligned}$$

式中 $\mathscr{X}$ 为局中人 Ⅰ 的纯策略集，$\mathscr{Y}$ 为局中人 Ⅱ 的纯策略集，$\boldsymbol{x}$ 和 $\boldsymbol{y}$ 分别为局中人 Ⅰ 和 Ⅱ 的混合策略，$\boldsymbol{x}\in\mathscr{X}$，$\boldsymbol{y}\in\mathscr{Y}$，$g_1(\boldsymbol{x})$ 和 $g_2(\boldsymbol{y})$ 分别是 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的概率密度函数，则称 $\overline{A}(x,y)$ 为此两人无限零和对策的赢得期望；称

$$A_1(\boldsymbol{x},\boldsymbol{y})=E^yA(\boldsymbol{x},\boldsymbol{y})=\int_{\mathscr{Y}}A(\boldsymbol{x},\boldsymbol{y})g_2(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$$

为局中人 Ⅰ 的赢得期望；称

$$A_2(\boldsymbol{x},\boldsymbol{y})=E^xA(\boldsymbol{x},\boldsymbol{y})=\int_{\mathscr{X}}A(\boldsymbol{x},\boldsymbol{y})g_1(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

为局中人 Ⅱ 的赢得期望.

混合平衡局势(mixed balance situation)　一类特殊的混合策略对. 设 $\overline{A}(\boldsymbol{x},\boldsymbol{y})$ 是两人无限零和对策的赢得期望，若存在混合策略 $\bar{\boldsymbol{x}},\bar{\boldsymbol{y}}$，使得对任意的 $\boldsymbol{x},\boldsymbol{y}$，有 $\overline{A}(\boldsymbol{x},\bar{\boldsymbol{y}})\leqslant\overline{A}(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}})\leqslant\overline{A}(\bar{\boldsymbol{x}},\boldsymbol{y})$，则称 $(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}})$ 是对策的混合平衡局势. 对于两人无限零和对策，若有 $(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}})$，使

$$\overline{A}(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}})=\sup_x\inf_y\overline{A}(\boldsymbol{x},\boldsymbol{y})=\inf_y\sup_x\overline{A}(\boldsymbol{x},\boldsymbol{y})$$

成立，则 $(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}})$ 是此对策的混合平衡局势. 设 $G=(\mathscr{X},\mathscr{Y};A)$ 是单位正方形上的连续对策，若 $\mathscr{X}=\mathscr{Y}\subset[0,1]\subset R$，且对策 G 的赢得函数 A 是 $\boldsymbol{x}\in\mathscr{X}$ 和 $\boldsymbol{y}\in\mathscr{Y}$ 的连续函数，则单位正方形上的连续零和对策的混合平衡局势存在，即 A 的最大最小值和最小最大值存在且相等.

凸对策(convex game)　一类有特殊赢得函数的对策. 如果对于局中人 Ⅰ 的任意纯策略 $\boldsymbol{x}\in\mathscr{X}$，赢得函数 $A(\boldsymbol{x},\boldsymbol{y})$ 是 $\boldsymbol{y}$ 的一个凸函数，称此对策 G 为凸对策. 凸对策的性质有：

1. 局中人 Ⅱ 必存在最优纯策略.

2. $V=\min\limits_{0\leqslant y\leqslant 1}\max\limits_{0\leqslant x\leqslant 1}A(\boldsymbol{x},\boldsymbol{y})=\max\limits_{0\leqslant x\leqslant 1}A(\boldsymbol{x},\bar{\boldsymbol{y}})$，

式中 $\bar{\boldsymbol{y}}$ 是局中人 Ⅱ 的最优纯策略.

3. 局中人 Ⅱ 的最优纯策略 $\bar{\boldsymbol{y}}$ 由方程

$$V=\max_{0\leqslant x\leqslant 1}A(\boldsymbol{x},\boldsymbol{y})$$

解得，但 $\boldsymbol{y}$ 的值不惟一. 若对策是严格凸对策(即对任意 $\boldsymbol{x}$，$A(\boldsymbol{x},\boldsymbol{y})$ 是 $\boldsymbol{y}$ 的严格凸函数)，则局中人 Ⅱ 的最优纯策略是惟一的.

凸对策的值是其赢得函数 $A(\boldsymbol{x},\boldsymbol{y})$ 在单位正方形一边 $\boldsymbol{x}\in[0,1]$ 上的最大值、在另一边 $\boldsymbol{y}\in[0,1]$ 上的最小值，即

$$V=\min_{0\leqslant y\leqslant 1}\max_{0\leqslant x\leqslant 1}A(\boldsymbol{x},\boldsymbol{y}).$$

凸对策的值(value of convex game)　见"凸对

策”.

可分对策(separable game)　一类特殊的无限对策. 其赢得函数(亦称可分函数)为

$$A(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} r_i(\boldsymbol{x}) s_j(\boldsymbol{y}),$$

式中 a_{ij} 是常数，$r_i(\boldsymbol{x})$ 和 $s_j(\boldsymbol{y})$ 是连续函数($i=1,2,\cdots,m$; $j=1,2,\cdots,n$). 当局中人 Ⅰ 期望支付的最小值等于局中人 Ⅱ 期望支付的最大值时，即

$$\min_{s\in S} E(\boldsymbol{r}^*,\boldsymbol{s}) = \max_{r\in R} E(\boldsymbol{r},\boldsymbol{s}^*) = E(\boldsymbol{r}^*,\boldsymbol{s}^*)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} r_i^* s_j^*$$

(式中 $\boldsymbol{r}^* = (r_1^*, r_2^*, \cdots, r_m^*) \in R$, $\boldsymbol{s}^* = (s_1^*, s_2^*, \cdots, s_n^*) \in S$) 时，则称这样的策略分别是局中人 Ⅰ 和 Ⅱ 在可分对策中的最优策略. 称这样的值 $E(\boldsymbol{r}^*,\boldsymbol{s}^*)$ 为可分对策的最优值.

可分函数(separable function)　见“可分对策”.

可分对策的最优策略(optimal strategy of separable game)　见“可分对策”.

矩空间(matrix space)　可分对策中分布函数的矩所构成的集合. 设可分对策的赢得函数是

$$A(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} r_i(\boldsymbol{x}) s_j(\boldsymbol{y}),$$

局中人 Ⅰ 的混合策略的分布函数是 $F(\boldsymbol{x})$，局中人 Ⅱ 的混合策略的分布函数是 $P(\boldsymbol{y})$，则局中人 Ⅰ 得到的期望支付是

$$E(r,s) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} r_i s_j,$$

式中

$$r_i = \int_0^1 r_i(\boldsymbol{x})\mathrm{d}F(\boldsymbol{x}) \quad (i = 1,2,\cdots,m),$$

$$s_j = \int_0^1 s_j(\boldsymbol{y})\mathrm{d}P(\boldsymbol{y}) \quad (j = 1,2,\cdots,n).$$

对于每一个分布函数 $F(\boldsymbol{x})$，有一个 r_i 与之对应，当 $F(\boldsymbol{x})$ 在全体分布函数集中变化时，点 r_i 的集合为 m 维欧氏空间的子集 R. 同样，对于每一个分布函数 $P(\boldsymbol{y})$，有一个 s_j 与之对应. 当 $P(\boldsymbol{y})$ 在全体分布函数集中变化时，点 s_j 的集合为 n 维欧氏空间的子集 S. 称集合 R 和 S 分别为函数 $\{r_i(x)\}$ 和 $\{s_j(y)\}$ 的矩空间.

不动点法(fixed point method)　亦称映射法. 对策问题的一类几何求解方法. 局中人 Ⅰ 的每一个最优策略是局中人 Ⅱ 每一个最优策略的像点. 同样，局中人 Ⅱ 的每一个最优策略是局中人 Ⅰ 的每一个最优策略的像点. 设 R^* 和 S^* 分别是局中人 Ⅰ 和 Ⅱ 的最优策略集. 若 $\boldsymbol{r}^* \in R^*$，则有

$$S(\boldsymbol{r}^*) = \{\boldsymbol{s} \mid E(\boldsymbol{r}^*,\boldsymbol{s}) = \min_{t\in S} E(\boldsymbol{r}^*,\boldsymbol{t})\} = S^*.$$

这就是说，$\boldsymbol{r}^*$ 的像点集是 S^*. 同理，若 $\boldsymbol{s}^* \in S^*$，则有

$$R(\boldsymbol{s}^*) = \{\boldsymbol{r} \mid E(\boldsymbol{r},\boldsymbol{s}^*) = \max_{t\in r} E(\boldsymbol{t},\boldsymbol{s}^*)\} = R^*.$$

这就是说，$\boldsymbol{s}^*$ 的像点集是 R^*. 最优策略 $\boldsymbol{r}^* \in R^*$ 和 $\boldsymbol{s}^* \in S^*$ 都是上述映射的不动点. 这样，求最优策略的问题就转化为求不动点的问题.

映射法(mapping method)　即“不动点法”.

定时对策(game of timing)　一类特殊的两人无限零和对策. 其赢得函数为

$$A(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} M_1(\boldsymbol{x},\boldsymbol{y}) & (x > y), \\ M_0(\boldsymbol{x},\boldsymbol{y}) & (x = y), \\ M_2(\boldsymbol{x},\boldsymbol{y}) & (x < y). \end{cases}$$

20 世纪 50 年代，曾用这类对策解决了一些军事问题，特别是建立了一些空军战略和空战的模型，得出了一些有意义的结果.

阵地对策(position game)　亦称展开型对策. 一种特殊对策. 是一种用树状图表示的对策. 树状图由顶点和连结顶点的线段——树枝组成. 顶点称为阵地，它表明对策中可能出现的状态. 如图所示，树状图有且仅有一个树根 x，树状图的悬挂点 q,r,s,t 称为边界点，它们分别代表对策可能出现的结局. 边界点旁的向量表示各局中人在此结局的所得. 对于树状图中的每一个非边界点，都将和某个局中人对应，如 y 点对应局中人乙，它表示乙在状态 y 可以做出下一步的选择. 若每个局中人在每次决定自己的下一步时，知道前面所有各步的结果，则此对策是全信息阵地对策. 对于全信息阵地对策，只要阵地(状态)总数是有限的，不论其是否为零和对策，也不论其有多少个局中人，则在非合作条件下必定有纯平衡局势，也就是有纯策略解. 阵地对策常用来描述完整的行动方案不易说清楚的对策，如下象棋，要走很多步，每一步又有多种走法，而且每一步走法要据前一步的结果来决定. 这类对策又可称为一步一步走的对策. 对策论的第一篇论文，即前面已经提到过的策梅洛(Zermelo, E. F. F.)所写的文章，就是用集合论的方法，研究了国际象棋的着法. 这类对策就是阵地对策，其精确定义是库恩(Kuhn, H. W.)给出的.

(−1,1)　(−1,1)　(−1,1)　(−1,1)
q　r　s　t
y 乙　z 乙
x,甲

展开型对策(extensive form game)　即“阵地对策”.

有序根树(ordered rooted tree)　一类特殊的有限集. 在有限集合 K 上定义了半序关系“$\precsim$”，它满足：

1. 存在元素 $\mathbf{0} \in K$，使得对任意的 $\boldsymbol{x} \in K$，都有

$\mathbf{0} \precsim x$.

2. 对于任意一对 $x, y \in K$，若存在 $z \in K$，使得

$$x \precsim z \text{ 及 } y \precsim z \Rightarrow x \precsim y \text{ 或 } y \precsim x,$$

则集合 K 称为有序根树，称 K 的顶点为阵地，元素 **0** 为原点或树根.

演化方程(evolution equation)　揭示动态对策变化规律的方程(参见"动态对策"). 即

$$z(k+1) = f_k(z(k), x(k), y(k)) \quad (k = 0, 1, \cdots, N-1).$$

这里考虑的是一个动态两人零和对策

$$G = \{\mathscr{X}_k, \mathscr{Y}_k; A_k\} \quad (k = 0, 1, \cdots, N-1),$$

式中 $\mathscr{X}_k = \{x(k)\}$，$\mathscr{Y}_k = \{y(k)\}$ 分别表示局中人 Ⅰ 和局中人 Ⅱ 在 k 时刻的策略集，A_k 是依赖于策略 $x(k)$ 和 $y(k)$ 的局中人 Ⅰ 的赢得函数，$z(k)$ 是状态变量，$f_k(\cdot)$ 是某个函数，当初态 $z(0)$ 已知时，局中人 Ⅰ 在整个对策过程的赢得为

$$A = A_N(z(N)) + \sum_{k=0}^{N-1} A_k(z(k), x(k), y(k)).$$

信息动态对策(information dynamic game)　一种动态对策. 指经典信息结构下的完全状态信息动态两人零和对策. 根据各局中人所获信息的多寡，此类对策可分为三种模式：

1. 开环模式. 指两个局中人在对策过程中只知道初始状态，而不管其后各时刻的状态. 信息结构可以表示为 $\eta(k) = \{z(0)\} = \eta^0(k)(k = 0, 1, \cdots, N-1)$，式中 $\eta(k)$ 为两个局中人在 k 时刻所获得的信息. 通常称在此信息结构式条件下得到的对策解为开环解.

2. 无记忆闭环模式. 此时，局中人在 k 时刻所获信息就是 k 时刻对策的状态，即对策的信息结构为 $\eta(k) = \{z(k)\} = \eta^j(k)$. 通常称在此信息结构式条件下得到的对策解为无记忆闭环解，或纯反馈解.

3. 闭环模式. 在闭环模式中，局中人在 k 时刻的信息不仅包含 k 时刻对策的状态，而且包含 k 以前的所有状态信息，即此时的信息结构是 $\eta(k) = \{z^k\} = \eta^c(k)$，式中 $z^k = (z(0), z(1), \cdots, z(k))$，此式表示各局中人拥有最大数据库，在此种信息结构条件下的对策解，称为闭环解.

对策律(game law)　信息动态对策中的对策规则. 两局中人的策略为

$$x(k) = r_{1k}(\eta(k)),$$
$$y(k) = r_{2k}(\eta(k)) \quad (k = 0, 1, \cdots, N-1),$$

称 $r_{ik}(\cdot)$ 为对策律 $(i = 1, 2)$. 对于三种不同的信息模式，对策律为

$$\begin{cases} \gamma_i^0 = (\gamma_{i0}^0, \gamma_{i1}^0, \cdots, \gamma_{iN-1}^0)^{\mathrm{T}}, \\ \gamma_i^j = (\gamma_{i0}^j, \gamma_{i1}^j, \cdots, \gamma_{iN-1}^j)^{\mathrm{T}}, \\ \gamma_i^c = (\gamma_{i0}^c, \gamma_{i1}^c, \cdots, \gamma_{iN-1}^c)^{\mathrm{T}} \quad (i = 1, 2), \end{cases}$$

$\gamma_i^0, \gamma_i^j, \gamma_i^c$ 构成的集合分别用 $\Gamma_i^0, \Gamma_i^j, \Gamma_i^c$ 表示.

非零和对策(non-zero-sum game)　一类非完全对策. 指所有局中人的支付之和不一定为零的对策，是既有对抗又有联合的缓和竞争的对策. 一般形式为

$$G = \{I, \{\alpha_i\}_{i \in I}, \{A_i\}_{i \in I}\} \quad \left(\sum_{i \in I} A_i \neq 0\right).$$

非零和对策与零和对策是有差异的. 零和对策是一种竞争、对抗的对策. 在现实世界中除了军事冲突或赌博之外，很少出现这种你死我活的局面. 非零和对策表现为既有对抗又有联合的缓和竞争.

非零和对策的混合策略(mixed strategy of non-zero-sum game)　非零和对策的纯策略的推广. 指局中人在概率意义下的策略. 常用于求某局势出现的概率. 每个局中人选取混合策略所构成的局势称为该非零和对策的混合局势. 用 $X_i (i \in I)$ 表示第 i 个局中人的混合策略，$X_i(\alpha_{ik_i})$ 表示 α_{ik_i} 的概率，并假定各局中人策略集上的概率分布互相独立. 设 $S = (\alpha_{1k_1}, \alpha_{2k_2}, \cdots, \alpha_{nk_n})$ 是对策的一个局势，$X(S)$ 表示此局势出现的概率，则

$$X(S) = X(\alpha_{1k_1}, \alpha_{2k_2}, \cdots, \alpha_{nk_n})$$
$$= X_1(S_{1k_1}) X_2(S_{2k_2}) \cdots X_n(S_{nk_n}).$$

非零和对策的赢得期望为

$$E_i(X) = \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} \cdots \sum_{k_n=1}^{m_n} \Big[A_i(\alpha_{1k_1}, \alpha_{2k_2}, \cdots, \alpha_{nk_n}) \cdot \prod_{j=1}^{n} X_j(\alpha_{jk_j}) \Big] \quad (i = 1, 2, \cdots, n).$$

非零和对策的赢得期望(win expectation of non-zero-sum game)　见"非零和对策的混合策略".

非零和对策的混合局势(mixed situation of non-zero-sum game)　见"非零和对策的混合策略".

非零和对策的平衡局势(balance situation of non-zero-sum game)　非零和对策的一种混合局势. 指策略集扩充为混合策略集后的平衡局势. 若 $\overline{X}$ 是非零和对策 G 的一个混合平衡局势，则有

$$A_i(\overline{X} \parallel x_i) \leqslant A_i(\overline{X}) \quad (i = 1, 2, \cdots, n),$$

式中 x_i 是局中人 i 的任一混合策略，$\overline{X} \parallel x_i$ 表示以 x_i 代替 $\overline{X}_i$ 后的一个混合局势. 利用混合策略寻求平衡局势是很困难的. 扩充后的有穷非零和对策必然存在平衡局势.

非合作型对策(non-cooperate form game)　一种对策. 是各局中人互不结成同盟的对策. 局中人只有两人时，称为两人非合作型对策. 局中人有三人或三人以上时，称为多人非合作型对策. 两人非合作型对策又称为双矩阵对策. 此时局中人的赢得矩阵用

双矩阵表示,即

$$C=(A,B)=\begin{bmatrix}(a_{11},b_{11}) & (a_{12},b_{12}) & \cdots & (a_{1n},b_{1n})\\(a_{21},b_{21}) & (a_{22},b_{22}) & \cdots & (a_{2n},b_{2n})\\\vdots & \vdots & & \vdots\\(a_{m1},b_{m1}) & (a_{m2},b_{m2}) & \cdots & (a_{mn},b_{mn})\end{bmatrix}.$$

用双矩阵表示局中人的混合策略称为双矩阵混合策略. 设局中人Ⅰ和Ⅱ的赢得分别是 $E_1(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{x}A\boldsymbol{y}^{\mathrm{T}}$ 和 $E_2(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{x}B\boldsymbol{y}^{\mathrm{T}}$,式中 $A=(a_{ij})_{m\times n}$, $B=(b_{ij})_{m\times n}$,则 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 分别是局中人Ⅰ和Ⅱ的混合策略.

不结盟对策(non-aligned game)　即"非合作型对策".

两人非合作型对策(two person non-cooperate form game)　见"非合作型对策".

双矩阵混合策略(mixed strategy of two matrix)　见"非合作型对策".

双矩阵对策的平衡局势(balance situation of two matrix game)　双矩阵对策中的一个特殊策略对. $(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}})$是双矩阵对策(两人非合作型对策)的平衡局势当且仅当

$$\begin{cases}A_{(i)}y^{\mathrm{T}}\leqslant \bar{x}A\bar{y}^{\mathrm{T}} & (i=1,2,\cdots,m),\\ xB^{(j)}\leqslant \bar{x}B\bar{y}^{\mathrm{T}} & (j=1,2,\cdots,n),\end{cases}$$

其中 $A_{(i)}$表示 A 的第 i 行向量$(i=1,2,\cdots,m)$, $B^{(j)}$表示 B 的第 j 列向量$(j=1,2,\cdots,n)$.

多人非合作型对策(many persons non-cooperate form game)　一种非合作型对策. 指具有两个以上局中人的非合作型对策(参见"非合作型对策"). n 人非合作型对策记为 $G(N,\{X^i\}_{i\in N},\{A_i\}_{i\in N})$,其中 $N=\{1,2,\cdots,n\}$表示局中人集合; $X^1,X^2,\cdots,X^n$ 表示局中人各自的策略集; $A_i(i=1,2,\cdots,n)$表示局中人 i 的支付函数, $A_i(\boldsymbol{x}^1,\boldsymbol{x}^2,\cdots,\boldsymbol{x}^n)$是指局中人1采用策略 $\boldsymbol{x}^1$,局中人2采用策略 $\boldsymbol{x}^2$, $\cdots$,局中人 n 采用策略 $\boldsymbol{x}^n$ 时,局中人 i 所得到的支付$(i=1,2,\cdots,n)$. 若对于每个 $i\in N$,有

$$A_i(\boldsymbol{x})=\max_{y^i\in X^i}A_i(\boldsymbol{x}\parallel \boldsymbol{y}^i)\quad(i=1,2,\cdots,n),$$

则称局势 $\boldsymbol{x}=(\boldsymbol{x}^1,\boldsymbol{x}^2,\cdots,\boldsymbol{x}^n)$为 G 的平衡局势,其中

$$\boldsymbol{x}\parallel \boldsymbol{y}^i=(\boldsymbol{x}^1,\boldsymbol{x}^2,\cdots,\boldsymbol{x}^{i-1},\boldsymbol{y}^i,\boldsymbol{x}^{i+1},\cdots,\boldsymbol{x}^n),$$

表示在局势 $\boldsymbol{x}$ 的基础上除局中人 i 选取了新策略 $\boldsymbol{y}^i\in X^i$ 外,其余局中人所取策略保持不变而得到的一个新局势. 在一定条件下,可以利用不动点定理证明多人非合作对策平衡局势的存在性.

合作型对策(cooperate form game)　一种对策. 各局中人之间可结成一定规模的同盟,决策之前进行必要的协商的对策称为合作型对策.

两人合作型对策中局中人Ⅰ和Ⅱ的赢得期望为

$$E_1=\sum_{\substack{1\leqslant i\leqslant m\\1\leqslant j\leqslant n}}P_{ij}a_{ij},\quad E_2=\sum_{\substack{1\leqslant i\leqslant m\\1\leqslant j\leqslant n}}P_{ij}b_{ij},$$

式中 m 和 n 分别表示局中人Ⅰ和Ⅱ有 m 种和 n 种纯策略,其赢得用双矩阵 $C=(A,B)$表示,其中 $A=(a_{ij})_{m\times n}$, $B=(b_{ij})_{m\times n}$; P_{ij}表示两局中人事先约定采用的纯策略组合$(\alpha_{i_1},\alpha_{i_2})$的概率,且

$$\sum_{\substack{1\leqslant i\leqslant m\\1\leqslant j\leqslant n}}P_{ij}=1.$$

当取不同概率分布时,此式确定的赢得期望集为一凸组合,其顶点即为不同纯策略组合下的赢得.

优势集(dominate set)　一种赢得集合. 指由不受其他赢得控制的赢得构成的集合. 设 E 是两人合作型对策的赢得期望集,而$(E_1,E_2)\in E$, $(E_1',E_2')\in E$ 是两个不同的赢得期望,且 $E_1>E_1'$, $E_2>E_2'$,则称(E_1,E_2)控制(E_1',E_2').

合同集(contract set)　一种赢得集合. 指优势集中赢得大于两局中人最大最小赢得的子集. 合作型对策的合同集一般含有无穷多个点,对求解对策是不便的. 下面给出惟一确定合作对策中最优策略的准则:令

$$(E_1',E_2')=G(E,(\overline{E}_1,\overline{E}_2)),$$

式中 G 是构造函数, E 是对策的赢得期望集, $(\overline{E}_1,\overline{E}_2)$是参考赢得,是两个局中人所能接受的最低赢得期望. 一般令

$$\overline{E}_1=\max_i\min_j a_{ij},\quad \overline{E}_2=\max_j\min_i b_{ij}$$

(通常称为现状点), (E_1',E_2')为按关系 G 求得的惟一赢得期望. 函数 G 须满足纳什公理(参见"纳什公理"),并且 $G(E,(\overline{E}_1,\overline{E}_2))$就是数学规划问题

$$\max(E_1-\overline{E}_1)(E_2-\overline{E}_2),\ (E_1,E_2)\in T$$

的最优解(E_1',E_2'),式中 T 是此对策的合同集. 合作型对策中,在合同集上求得的最优解称为合同解.

现状点(present situation point)　见"合同集".

合同解(contract solution)　见"合同集".

纳什公理(Nash axiom)　合作型对策有最优解的公设条件. 它们是:

1. $(E_1',E_2')\in E$,且 $E_1'\geqslant\overline{E}_1$, $E_2'\geqslant\overline{E}_2$.

2. 设 T 是平面 $E_1\sim E_2$ 上的某个线性函数,亦即 $T(E_1,E_2)=(a_1E_1+b_1,a_2E_2+b_2)$(式中 a_1,a_2,b_1,b_2 均为常数),则 $G(T(E),T(\overline{E}_1,\overline{E}_2))=T(E_1',E_2')$,即 G 和 T 的作用次序可以互换.

3. 设 $\widetilde{E}\subset E$ 是子集,且$(\overline{E}_1,\overline{E}_2)\in\widetilde{E}$, $(E_1',E_2')\in\widetilde{E}$,则 $G(\widetilde{E},(\widetilde{E}_1,\widetilde{E}_2))=G(E,(\overline{E}_1,\overline{E}_2))$.

4. 若由$(E_1,E_2)\in E$ 可推知$(E_2,E_1)\in E$,且 $\overline{E}_1=\overline{E}_2$,则有 $E_1'=E_2'$.

(此条目中的字母的意义参考"优势集"和"合同集").

威胁解(menace solution)　对策问题的一类解.是按照特定准则求解所得到的策略对.局中人Ⅰ选择混合策略 $\boldsymbol{x}$,使自己的赢得期望尽可能大,而使对方的赢得期望尽可能小;而局中人Ⅱ选择混合策略 $\boldsymbol{y}$,使自己的赢得期望尽可能大,而使对方的赢得期望尽可能小,称此类解为威胁解.在威胁解中,双方都企图通过谈判选择适当的策略作为现状点的策略对,使求得的最优赢得期望对己方有利,对对方不利.在求威胁解时,不把固定的最大最小收入作为函数 G 的现状点,而代之以某个赢得期望 $(\overline{E}_1,\overline{E}_2)=(\boldsymbol{x}A\boldsymbol{y}^{\mathrm{T}},\boldsymbol{x}B\boldsymbol{y}^{\mathrm{T}})$,式中 A,B 分别是局中人Ⅰ和局中人Ⅱ的赢得矩阵,$(\boldsymbol{x},\boldsymbol{y})$ 是策略对,按此式求解理想的赢得期望为

$$(E_1',E_2')=G(E,(\boldsymbol{x}A\boldsymbol{y}^{\mathrm{T}},\boldsymbol{x}B\boldsymbol{y}^{\mathrm{T}})),$$

赢得集 E 不变,E_1',E_2' 则为策略对的函数,即

$$E_1'=G_1(\boldsymbol{x},\boldsymbol{y}),\quad E_2'=G_2(\boldsymbol{x},\boldsymbol{y}).$$

多人合作型对策(many person cooperate form game)　一种合作型对策.是允许局中人结盟的多人对策.在这种对策中,局中人之间可能结成几个同盟,在同盟内部也可能采取不同的赢得分配方式.设对策 $G=(I,\{S_i\}_{i\in I},\{A_i\}_{i\in I})$,$I=\{1,2,\cdots,n\}$ 是局中人集,其中某些子集 $\tilde{I}$ 的局中人结成同盟;在整个对策过程中,$\tilde{I}$ 中的局中人将以一个整体出现,每个局中人都力图使本同盟的总赢得为最大,这就构成了多人合作对策.当局中人结成同盟 $\tilde{I}$ 时,同盟的策略集定义为

$$S_{\tilde{I}}=\prod_{i\in\tilde{I}}S_i,$$

同盟的赢得定义为其所有局中人赢得之和:

$$A_{\tilde{I}}(S)=\sum_{i\in\tilde{I}}A_i(S),$$

式中 $S=(S_1,S_2,\cdots,S_n)$ 是对策的一个局势.

多人合作型对策中,同盟 $\tilde{I}$ 所能保证得到的最大收入称为同盟 $\tilde{I}$ 的特征函数,以 $V(\tilde{I})$ 表示,如

$$V(\tilde{I})=\max_{X\in S_{\tilde{I}}}\min_{Y\in S_{I\setminus\tilde{I}}}A_{\tilde{I}}(X,Y).$$

以同盟 $\tilde{I}$ 为一方,以局中人集合中除 $\tilde{I}$ 的局中人外其余局中人构成的同盟 $I\setminus\tilde{I}$ 为另一方,按赢得函数 $A_{\tilde{I}}$ 形成一个两人零和对策 $(S_{\tilde{I}},S_{I\setminus\tilde{I}};A_{\tilde{I}})$.同盟 $\tilde{I}$ 的最大最小赢得即同盟 $\tilde{I}$ 所能得到的最大赢得.若同盟 $\tilde{I}=\varnothing$,则 $V(\varnothing)=0$.若 V_1,V_2 都是定义在局中人集 I 上的特征函数,任取 $\lambda_1\geqslant0,\lambda_2\geqslant0$,则 $V=\lambda_1V_1+\lambda_2V_2$ 也是定义在 I 上的特征函数.有一类特殊的合作对策,称为零对策,其特征函数恒等于零.称满足

$$\sum_{i\in I}V(i)=V(I)$$

的对策为非本质合作对策.否则,称为本质合作对策.

同盟的特征函数(characteristic function of cooperate $\tilde{I}$)　见"多人合作型对策".

非本质合作对策(non-essential cooperate game)　见"多人合作型对策".

本质合作对策(essential cooperate game)　见"多人合作型对策".

策略等价(strategic equivalence)　策略对之间的关系.设对策 G 具有局中人集合 I 和定义在其上的特征函数 V,另一合作对策 G' 也具有局中人集合 I,但定义在其上的特征函数是 V'.若存在正实数 k 和任意实数 $C_i(i\in I)$,使得对任意同盟 $\tilde{I}\subset I$ 有等式

$$V'(\tilde{I})=kV(\tilde{I})+\sum_{i\in\tilde{I}}C_i$$

成立,则称 G 和 G' 是策略等价的,并称 V 和 V' 等价,记为 $V\sim V'$.任一非本质对策的策略都等价于某个零对策.

(0,1)规范化((0,1) standardization)　对策论的基本概念之一.对策 G 和特征函数 V 为(0,1)规范化是指,n 人合作对策 $G\equiv[I,V]$ 的特征函数 V 满足下面两个条件:$V\{i\}=0\quad(i=1,2,\cdots,n)$;$V(I)=1$.每一个实质性的 n 人合作对策 $G\equiv[I,V]$ 策略等价于惟一的一个(0,1)规范化对策.

常和合作对策(constant-sum cooperate game)　一类合作对策.设 $V(\tilde{I})+V(I\setminus\tilde{I})=V=$ 常数 $(\forall\tilde{I}\subset I)$,式中 $\tilde{I}\subset I$ 是对策中的一个同盟,$I\setminus\tilde{I}$ 是该对策的另一个同盟,则称此对策为常和合作对策.仅有两个局中人的常和合作对策必是非本质合作对策.有三个局中人的常和合作对策可分为非本质合作对策等价类和本质合作对策等价类两大类.在一般情况下,若对策属常和对策,而对策的局中人数大于3,则策略等价类将大大增加.

合作对策的赢得分成(win share of cooperate game)　一类对策赢得与特征函数之间的关系.满足

$$\begin{cases}x_i\geqslant V(i),\quad\forall i\in I\quad(\text{个体合理性}),\\ \sum_{i\in I}x_i=V(I)\quad(\text{总体合理性})\end{cases}$$

的向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$ 称为对策的赢得分成,式中 x_i 是局中人 i 最终得到的赢得,$I=\{1,2,\cdots,n\}$ 是合作对策的局中人集合,V 是定义在局中人集合 I 上的特征函数.一个非本质合作对策仅有一个分成,即 $\boldsymbol{x}=(V(1),V(2),\cdots,V(n))$.任何至少有两个局中人的本质合作对策都有无穷多个分成,构成局中人的分成集.

经典合作对策(classical cooperate game)　一类按分成决定的对策.是具有局中人集合 I 和定义

在其上的特征函数 V，一组分成 $\boldsymbol{x}$ 满足

$$\begin{cases} x_i \geqslant V(i) \quad (\forall i \in I), \\ \sum_{i \in I} x_i = V(I) \end{cases}$$

的合作对策. 向量 $\boldsymbol{x}$ 是经典合作对策的一个分成的充分必要条件是

$$x_i = V(i) + a_i \quad (\forall i \in I),$$

式中

$$a_i \geqslant 0 \quad (\forall i \in I), \sum_{i=1}^{n} a_i = V(I) - \sum_{i=1}^{n} V(i).$$

转归(individual rationality) 亦称支付或分配. 合作对策术语. 指一类特殊赢得的组合. 转归用一个 n 维向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$ 表示，其中 x_i 是局中人 i 在赢得组合(即转归)$\boldsymbol{x}$ 中应得的份额. 向量 $\boldsymbol{x}$ 应满足条件：

1. $x_i \geqslant V(\{i\}) \quad (i=1,2,\cdots,n)$.

2. $\sum_{i=1}^{n} x_i = V(I)$.

非实质性对策只有一个转归，即 $x=(V\{1\},V\{2\},\cdots,V\{n\})$. 实质性的对策 $G\equiv[I,V]$有无穷多个转归.

支付(payoff) 即“转归”.

分配(allocate) 即“转归”.

个体合理性条件(individual rationality condition) 合作对策的前提条件. 指局中人 i 在赢得组合(即转归)$\boldsymbol{x}$ 中应得份额所满足的条件，即

$$x_i \geqslant V(\{i\}) \quad (i=1,2,\cdots,n)$$

(参见“转归”). 这是因为，一个局中人 i，如果参加某个联盟得到分配的份额 x_i 还不如个人单干得到的收入，这样的分配是不会被接受的.

集体合理性条件(group rationality condition) 亦称帕雷托最优性条件. 合作对策的最优性条件. 即

$$V(I) > \sum_{i=1}^{n} x_i.$$

该条件表明联盟的总收入为 $V(I)$，每个局中人 i 在 x_i 外，还可得到一些额外的收入，所以，I 中全部成员可以组成联盟.

帕雷托最优性条件(Pareto optimality condition) 即“集体合理性条件”.

优超域(dominates dominion) 合作对策的基本概念之一. 指转归关于联盟的优超集合. 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$和 $\boldsymbol{y}=(y_1,y_2,\cdots,y_n)$是 n 人合作对策 $G\equiv[I,V]$的两个转归，联盟 $\tilde{I}\subset I$ 非空. 若

$$V(\tilde{I}) \geqslant \sum_{i \in \tilde{I}} y_i \text{ 且 } y_i > x_i \quad (i \in I),$$

则称 $\boldsymbol{y}$ 关于 $\tilde{I}$ 优超于 $\boldsymbol{x}$，记为 $\boldsymbol{y} \underset{\tilde{I}}{\succ} \boldsymbol{x}$. 集合

$$\mathrm{Dom}_{\tilde{I}}\, y = \{x \mid \boldsymbol{y} \underset{\tilde{I}}{\succ} \boldsymbol{x}\}$$

是转归 $\boldsymbol{y}$ 关于联盟 $\tilde{I}$ 的优超域.

对策的核心(core of game) 合作对策的基本概念之一. 指一个多人合作对策的赢得分成集的某个子集. 设 $G\equiv[I,V]$是 n 人合作对策，对于一切联盟 $\tilde{I}\subseteq I$，满足 $V(\tilde{I})\leqslant \boldsymbol{x}(\tilde{I})$的转归 $\boldsymbol{x}$ 组成的集合 C 称为 G 的核心，其中

$$\boldsymbol{x}(\tilde{I}) = \sum_{i \in \tilde{I}} x_i.$$

即对策 G 的核心 $C=\{\boldsymbol{x} \mid \boldsymbol{x}\in X; V(\tilde{I})-\boldsymbol{x}(\tilde{I})\leqslant 0, \tilde{I}\subset I\}$. $V(\tilde{I})\leqslant \boldsymbol{x}(\tilde{I})$表示：对于一切联盟 $\tilde{I}\subseteq I$，C 中的转归 $\boldsymbol{x}$ 提供给 $\tilde{I}$ 的分配不少于自身得到的总收入 $V(\tilde{I})$. 若 $G\equiv[I,V]$是 n 人合作对策，则转归 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)\in C$ 的充分必要条件是 $\boldsymbol{x}$ 不被优超. 若 $G\equiv[I,V]$是实质性的常和 n 人合作对策，则对策的核心 $C=\varnothing$.

诺伊曼-莫根施特恩解(Neumann-Morgenstern solution) 亦称稳定集. 合作对策的基本概念之一. 指合作对策分成集的某个子集. 冯 · 诺伊曼(von Neumann, J.)和莫根施特恩(Morgenstern, O.)曾经提出把满足所谓内部稳定性和外部稳定性的一组分成作为合作对策的解. 但这样的解对于本质合作对策并不惟一，甚至还比对策核中(参见“对策的核Ⅰ”、“对策的核Ⅱ”)的元素多. 内部稳定性指集合 R 中的分成不优于集内其他分成. 外部稳定性指对集合 R 以外的任意分成 $\boldsymbol{x}$，存在 R 内的分成 $\boldsymbol{y}$ 优于 $\boldsymbol{x}$. 若一个合作对策具有核 C 和诺伊曼-莫根施特恩解 R，则 C 是 R 的子集. 若一个合作对策的诺伊曼-莫根施特恩解 R 仅由惟一的一个分成 $\boldsymbol{x}$ 构成，则此对策的特征函数 V 是非本质的. 任何非本质合作对策的诺伊曼-莫根施特恩解为

$$\boldsymbol{x}=(V(1),V(2),\cdots,V(n)).$$

但对于具有两个或两个以上局中人的本质合作对策，诺伊曼-莫根施特恩解一般不止一个.

超出值(excess value) 合作对策术语. 指合作对策的联盟 S 关于转归向量 $\boldsymbol{x}$ 的超出值，记为 $e(S,\boldsymbol{x})$. 设合作对策 G 的局中人集合为 I，$S\subset I$ 为局中人的任一联盟，V 为 G 的特征函数，则称 $V(S)-\boldsymbol{x}(S)$为联盟 S 关于支付向量的超出值. 令 $T_{ij}=\{S\subset I \mid i\in S, j\notin S\}$，称

$$s_{ij}(\boldsymbol{x}) = \max_{S\in T_{ij}} e(S,\boldsymbol{x})$$

为在 $\boldsymbol{x}$ 处局中人 i 超过 j 的最大超出值. 若 $s_{ij}(\boldsymbol{x})>s_{ji}(\boldsymbol{x})$且 $x_j>V(j)$，则称 i 关于 $\boldsymbol{x}$ 胜过 j. 若两个局中人 i 和 j 中的任何一个都不胜过另一个(关于 $\boldsymbol{x}$)，则称 i 和 j 关于 $\boldsymbol{x}$ 处于平衡状态.

对策的核Ⅰ(core Ⅰ of game) 合作对策的基本概念之一. 指合作对策转归 $\boldsymbol{x}$ 的集合，记为 K. 在 K 中的每一个 $\boldsymbol{x}$ 处，每两个局中人 i,j 关于 $\boldsymbol{x}$ 都处于平衡状态.

联盟结构(coalition structure) 合作对策的基本概念之一.指全体局中人的集合 I 的一个划分 B,即 $B=\{B_1,B_2,\cdots,B_n\}$,式中 B_k 互不相交且

$$\bigcup_{k=1}^{n} B_k=I.$$

个体合理支付构形(individual rational payoff configuration) 合作对策的基本概念之一.指由支付向量和联盟结构向量为分量的向量.即

$$(\boldsymbol{x};B)\equiv(x_1,x_2,\cdots,x_m;B_1,B_2,\cdots,B_n).$$

式中 B 是一个联盟结构,$\boldsymbol{x}=(x_1,x_2,\cdots,x_m)$ 是一个支付向量,且满足

$$x_i\geqslant V(i)\quad(i=1,2,\cdots,m),$$
$$x(B_k)=V(B_k)\quad(k=1,2,\cdots,n).$$

对策的核Ⅱ(core Ⅱ of game) 一类支付构形集.指对策 G 关于联盟结构 B 的核 K 是一切个体合理支付构形 $(\boldsymbol{x};B)$ 的集合.设 B 是 n 人合作对策 $G\equiv[I,V]$ 的联盟结构.G 关于 B 的核 K 是一切个体合理支付构形 $(\boldsymbol{x};B)$ 的集合;每一对属于同一个 $B_k\in B$ 的局中人 i,j 在 $(\boldsymbol{x};B)$ 处于平衡状态.

对策的核仁(nucleus of game) 合作对策的基本概念之一.指 n 人合作对策 $G\equiv[I,V]$ 的核仁是转归 $\boldsymbol{x}\in X$ 的集合 N.在其中每一个 $\boldsymbol{x}$ 处,$\boldsymbol{\theta}$ 在字典次序上为最小,即 $N=\{\boldsymbol{x}\mid \boldsymbol{x}\in X$;对于一切 $\boldsymbol{y}\in X$ 有

$$\boldsymbol{\theta}(\boldsymbol{x})\underset{L}{\precsim}\boldsymbol{\theta}(\boldsymbol{y})\},$$

其中 $\boldsymbol{\theta}(\boldsymbol{x})$ 为 2^n 维向量,其分量是 I 的诸子集关于 $\boldsymbol{x}$ 的超出值,且 $\boldsymbol{\theta}(\boldsymbol{x})$ 的分量是按从大到小顺序排列的($\underset{L}{\precsim}$ 是字典序).核仁的主要性质有:

1. n 人合作对策 $G\equiv[I,V]$ 的核仁非空.

2. n 人合作对策 $G\equiv[I,V]$ 的核仁是 N,则 $|N|=1$,即由惟一的一个转归构成.

3. 设 n 人合作对策 $G\equiv[I,V]$ 的核是 K,核仁是 N,则 $N\subset K$.

支柱(carrier) 局中人组成的一类集合.若对于一切联盟 $\bar{I}\subset I$ 有 $V(\bar{I})=V(\bar{I}\cap N)$,式中 $N\subseteq I$ 是一个联盟,则称 N 为对策的支柱.若 N 是对策 V 的一个支柱,则包含 N 的任何集合仍是 V 的支柱.

夏普利值(Shapley value) 一类支付方案.是一种将 $V(I)$ 分配给 n 个局中人的支付方案.n 人合作对策 $\Gamma=\{I,V\}$ 的夏普利值是定义在 V 上的 n 维向量

$$\boldsymbol{\varphi}(V)=(\mathscr{S}_1(V),\mathscr{S}_2(V),\cdots,\mathscr{S}_n(V)),$$

它满足如下三个条件:

1. 对称性.若存在 I 的一个排列 $\pi\in\pi(I)$,对每个 $S\subseteq I$ 有 $V(\pi S)=V(S)$,则 $\mathscr{S}_{\pi i}(V)=\mathscr{S}_i(V)$.

2. 有效性.对于 Γ 的每一个支柱 N,都有

$$\sum_{i\in N}\mathscr{S}_i(V)=V(N).$$

3. 聚合性.对于定义在 I 的全体子集类上的任意两个特征函数 V 和 W,都有

$$\boldsymbol{\varphi}(V+W)=\boldsymbol{\varphi}(V)+\boldsymbol{\varphi}(W),$$

其中排列 π 是 I 到自身的一个一一映射,用 $\pi i=j$ 表示 j 是 i 在 π 下的象,πS 表示联盟 S 在 π 下的象集,$\pi(I)$ 表示定义在 I 上的全体排列组成的集合.1952 年,夏普利(Shapley,L. S.)根据对称性、有效性和聚合性三条公理定义了 n 人合作对策的全体局中人的一组值,而且满足这三条公理的值是惟一的.n 人合作对策 $G\equiv[I,V]$ 的夏普利值是

$$\Phi(V)=(\Phi_1(V),\cdots,\Phi_n(V)),$$
$$\Phi_i(V)=\sum_{\bar{I}(i)}\frac{(n-|\bar{I}|)!(|\bar{I}|-1)!}{n!}\cdot[V(\bar{I})-V(\bar{I}|i)]\quad(i=1,2,\cdots,n),$$

其中 $\bar{I}(i)$ 及 $\bar{I}$ 表示含局中人 i 的同盟.

递阶对策(hierarchical game) 一种对策.是具有等级差异的局中人的对策.它研究参与对策的各个局中人在经济活动中处于不同的“权力”地位上,形成一种层次结构(递阶结构)的一类动态对策问题的对策理论.具有支配全局或局部能力的局中人称为领导,其他根据领导的策略选择自己的策略的局中人称为随从.根据对策的性质大致分为合作型、非合作型和对峙型三类.

斯塔科尔堡对策(Stackelberg game) 即“递阶对策”.

静态递阶对策(quiescent hierarchical game) 一种递阶对策.即单段递阶对策.设局中人Ⅰ和Ⅱ分别为领导和随从,U_1 和 U_2 分别为Ⅰ和Ⅱ的策略集,Ⅰ和Ⅱ的目标函数(损失函数)分别为 J_1 和 J_2,且

$$\begin{cases}J_1=J_1(u_1,u_2),\\J_2=J_2(u_1,u_2),\end{cases}$$

$u_1\in U_1,u_2\in U_2$.当领导采取策略 $\bar{u}_1\in U_1$,随从采取策略 $\bar{u}_2\in U_2$,且使 $J_2(\bar{u}_1,\bar{u}_2)\leqslant J_2(\bar{u}_1,u_2)$ $(\forall u_2\in U_2)$,则称 $\bar{u}_2$ 为随从对领导的策略 $\bar{u}_1$ 的合理反应.若存在定义在 U_1 上的某个函数 T,对任意的 $u_1\in U_1,u_2=T(u_1)\in U_2$ 是随从的合理反应,则

$$J_2(u_1,T(u_1))\leqslant J_2(u_1,u_2)\quad(\forall u_1\in U_1,\forall u_2\in U_2).$$

称集合 $R=\{u_2\in U_2,u_2=T(u_1),\forall u_1\in U_1\}$ 为随从对领导的合理反应集.若存在策略对 $(u_{1S},u_{2S})\in U_1\times U_2$,使 $J_1(u_{1S},u_{2S})\leqslant J_1(u_1,T(u_1))$ $(\forall u_1\in U_1)$,则称 (u_{1S},u_{2S}) 为对策的递阶平衡解,其中 $u_{2S}\in R$,即 $u_{2S}=T(u_{1S})$.

多步对策(multistep game) 一种对策.指局中人的每个策略须由多次行动实现的对策.在多步对策中,每个局中人的策略由一系列行动所组成.通常,局中人最初选择的只是他的一个策略中的某一具体行动,在其他局中人选择了他们的策略中的某

一行动后，这个局中人再采取新的行动……，直到对策结束. 例如，象棋、桥牌比赛等均可视为多步对策. 阵地对策是一类特殊的多步对策.

随机对策(stochastic game) 一类带有随机因素的多步对策. 给定 N 个基本对策 $G_k(k=1,2,\cdots,N)$，G_k 可用一个 $m_k\times n_k$ 阶支付矩阵 A_k 来表示，是两人零和对策，$A_k=(a_{ij}^k)_{m_k\times n_k}$. 这个多步对策是：从某一个基本对策开始，当对策处于第 k 个基本对策 G_k 时，若局中人 Ⅰ 和局中人 Ⅱ 分别选取纯策略 α_i 和 β_j，则局中人 Ⅰ 获得 a_{ij}^k，局中人 Ⅱ 付出 a_{ij}^k，并且以概率 $s_{ij}^k>0$ 停止整个对策，以概率 q_{ij}^{kl} 转移到基本对策 G_l 上去，直到对策结束. 对任意的 i,j,k，都有

$$s_{ij}^k+\sum_{l=1}^{N}q_{ij}^{kl}=1.$$

微分对策(differential game) 一种对策. 指应用微分方程组来描述连续动态决策过程的对策. 微分对策起源于研究军事上的追踪问题，但它不是单纯的追和逃的追踪问题. 微分对策是在 20 世纪 50 年代中期发展起来的一种新的对策理论，已被广泛应用于最优控制、系统工程和军事决策中. 在微分对策中，控制规律表示为状态变量的函数，向量 $\boldsymbol{x}(t)$ 表示时刻 t 时双方的状态变量，$\varphi(\boldsymbol{x})$ 表示追方的控制量，$\psi(\boldsymbol{x})$ 表示逃方的控制量. 当 $\varphi(\boldsymbol{x})$ 和 $\psi(\boldsymbol{x})$ 选定后，双方的运动情况可由微分方程 $d\boldsymbol{x}/dt=f(\boldsymbol{x},\varphi,\psi)$ 来描述，当给定初始状态 $\boldsymbol{x}^0$ 后，各方选择控制规律 $\varphi(\boldsymbol{x})$ 和 $\psi(\boldsymbol{x})$，即可求出在这局对策中自始至终的运动路线.

决 策 论

决策论(decision making theory) 运筹学的一个分支. 研究依据问题的属性，从多个可供选择的方案中如何选取最好或满意方案的学科. 决策是人们生活和工作中普遍存在的一种活动，小至个人生活、企业经营管理，大至国家的经济、政治等大政方针，随时都需要决策. 一般地，决策分为确定型、风险型和不确定型三类. 确定型决策又分为静态确定型和动态确定型两种；不确定型决策分为静态不确定型和动态不确定型两种. 而风险型和不确定型等决策问题，都是随机型决策问题. 随机型决策的基本特点是：后果的不确定性和后果的效用表示. 如果决策者采用的策略和依据的客观条件(简称状态)是不确定的，那么决策者做出某种决策以后，出现的后果也将会是不确定的. 后果的这种不确定性是随机型决策问题的主要特征之一. 另一个特点是：在进行决策之前，必须确定各种后果的效用，效用是对后果价值的定量分析.

在决策论中，分析含有不确定性的决策问题的理论和方法部分，称为决策分析. 1931 年，拉姆齐(Ramsey，F. P.)根据主观概率和效用这两个概念提出了制定决策的理论. 1937 年，德・菲莫蒂(De Fimetti，B.)对主观概率的结构研究做出了贡献. 1944 年，冯・诺伊曼(von Neumann，J.)和莫根施特恩(Morgenstern，O.)制定了不确定性决策问题的公理系，建立了现代效用理论. 在 20 世纪 50 年代初，瓦尔德(Wald，A.)和沙万奇(Savage，L. J.)提出了统计决策问题，并建立了理论体系. 兰法(Raiffa，H.)和沙莱佛(Schlaiffer，R. O.)等在 20 世纪 60 年代进一步发展了统计决策理论，形成了贝叶斯分析. 1966 年，霍华德(Howard，R. A.)把系统分析方法和统计决策理论结合起来，形成了决策分析. 1977 年，贝尔(Bell，D. E.)、克尼(Keeny，R. L.)和兰法认为，“决策分析的目的，是将一个问题分成两个方面：一方面表示每个方案的不同结果的可能性；另一方面评估这些结果的理想程度”.

由于社会的发展，使得要考虑的决策问题的因素愈来愈复杂，涉及的目标愈来愈多. 于是，自 20 世纪 70 年代以来，关于多目标(多指标，多属性)决策(multiple criteria decision making，简记为 MCDM)的理论和应用研究取得了重大进展. 目前，多目标决策(MCDM)已发展成为运筹学的一个独立的活跃分支.

决策系统(decision system) 决策论所研究的系统. 一般由输入、输出和内部结构三部分组成. 它的数学模型描述研究对象的状态演化方程和输出方程，体现决策者所需求的目标. 这样，只有在制定合理的目标函数之后，才有可能求得最优决策. 经典系统理论着重研究系统的输入输出关系，不涉及系统的内部结构和内部行为，而现代系统理论则同时注重两个方面的研究. 在系统理论中，输入输出关系常用传递函数来描述，系统的内部行为常用状态变量来描述，并以状态变量为中介把输入和输出联系起来. 这里的状态变量必须具有两个特性：

1. 能全面概括系统在 t 时刻以前和 t 时刻内系统运动的全部信息.

2. 根据系统现在的状态和从现在起到以后任何时刻间的输入，应能预测或完全确定那个时刻的输出，而且状态本身也得到发展.

不确定性系统(nondeterministic system) 亦称随机系统. 含有随机因素的决策系统. 设 $\boldsymbol{x}(t)=(x_1(t),x_2(t),\cdots,x_n(t))^{\mathrm{T}}$ 表示系统的状态向量，$\boldsymbol{y}(t)=(y_1(t),y_2(t),\cdots,y_m(t))^{\mathrm{T}}$ 表示系统的输出向量，$\boldsymbol{u}(t)=(u_1(t),u_2(t),\cdots,u_r(t))^{\mathrm{T}}$ 表示系统的输入向量(或控制向量，决策向量)，$\boldsymbol{m}(0)=(m_1(0),m_2(0),$

$\cdots, m_n(0))^{\mathrm{T}}$表示系统的初始状态，则第 i 个状态变量的演化方程为 $x_i(t)=f_i(t;\boldsymbol{x}(t-1),\boldsymbol{u}(t-1))+\xi_i(t)(i=1,2,\cdots,n;t=1,2,\cdots,N;\xi_i(t)(i=1,2,\cdots,n)$表示误差项)，初始状态为 $x_i(0)=m_i(0)(i=1,2,\cdots,n)$；输出(或量测)方程为 $y_j(t)=h_j(t;\boldsymbol{x}(t))+\eta_j(t)(j=1,2,\cdots,m;t=1,2,\cdots,N;h_j(\cdot)$表示第 j 个输出变量的输出函数，$\eta_j(t)(j=1,2,\cdots,m)$表示误差项).

随机系统(stochastic system)　即"不确定性系统".

确定性系统(deterministic system)　不含随机因素的决策系统.

线性系统(linear system)　一种决策系统. 指状态变量的演化函数和输出函数都是线性的系统. 即系统的状态向量 $\boldsymbol{x}(t)=A(t)\boldsymbol{x}(t-1)+B(t)\boldsymbol{u}(t-1)+\xi(t)(t=1,2,\cdots,N)$，$\boldsymbol{x}(0)=m(0)$，系统的输出向量$\boldsymbol{y}(t)=C(t)\boldsymbol{x}(t)+\eta(t)(t=1,2,\cdots,N)$，式中 $A(t)=(a_{ij}(t))_{n\times n}$，$B(t)=(b_{ij}(t))_{n\times r}$，$C(t)=(c_{ij}(t))_{m\times n}$ 都是矩阵. 称此系统为线性系统. 否则，称为非线性系统.

非线性系统(nonlinear system)　见"线性系统".

时不变系统(time-invariant system)　亦称定常系统. 一种决策系统. 指不依时间而改变的决策系统. 即状态变量的演化函数 $f(t;\cdot)$ 和输出变量的输出函数 $h(t;\cdot)$ 都不依赖于时间参数 t 的系统. 否则，称为时变系统.

时变系统(time-varying system)　见"时不变系统".

决策律(decision law)　用函数表示决策的方法. 把决策(或策略)表示为抽样试验结果 x 的某个函数 $\gamma(x)$ 时，称 $\gamma(\cdot)$ 为此决策问题的决策律. 如果 $\gamma(x)$是 x 的线性函数，称为线性决策律，否则称为非线性决策律.

风险矩阵(risk matrix)　亦称益损矩阵. 一种特殊矩阵. 是由风险值或益损值构成的矩阵.

决策模型(decision model)　决策问题的模型. 即决策问题的数学表示式.

确定型决策问题(deterministic type decision problem)　一类决策问题. 是把确定的自然状态看做必然事件(即其出现的概率为 1)，其他自然状态看做不可能事件(其出现的概率为 0)，并且符合下列四个条件的决策问题：

1. 存在决策者期望达到的一个明确目标.
2. 只存在一个确定的自然状态.
3. 存在着可供决策者选择的两个或两个以上的行动方案.
4. 不同的行动方案在确定状态下的益损值可以计算出来.

风险型决策问题(risk type decision problem)　亦称统计型决策问题或随机型决策问题. 一类决策问题. 是含有随机因素的决策问题. 它必须具备下列五个条件：

1. 具有决策者期望达到的目标.
2. 至少有两个以上的行动方案可供决策者选择，最后只选定一个满意的方案.
3. 存在两个或两个以上的自然状态.
4. 不同的行动方案在不同的自然状态下的益损值都可以计算出来.
5. 各种不同的自然状态出现的可能性(概率)可以预先估计或计算出来.

最大可能准则(maximal feasible criterion)　一种判别决策优劣的标准. 是在风险型决策问题中，选择一个可能性最大(概率最大)的自然状态进行决策，而不考虑其他自然状态. 它把风险型决策问题化成了确定型决策问题.

最大期望收益准则(expected monetary value)　简记为 EMV. 一种判别决策优劣的标准. 指在风险型决策问题中，以收益的期望值最大为决策的判别准则.

最小机会损失准则(expected opportunity loss)　简称 EOL. 一种判别决策优劣的标准. 在风险型决策问题中，以损失的期望值最小为决策优劣的判别准则.

决策树法(decision tree method)　一种用图示进行的决策方法. 决策思路如树枝形状，在做出决策之前要考虑各种可能发生的情况，还要预测未来的变化发展情况. 一般地，决策树法按如下步骤进行决策：

1. 画决策树.
2. 预计可能事件发生的概率，并把概率值标在树图的相应位置上.
3. 计算益损期望值，并根据益损期望值，取其极大值或极小值.

多级决策问题(multilevel decision problem)　一类决策问题. 是具有两级和两级以上的决策问题. 只包括一级的决策问题，称为单级决策问题. 对多级决策问题，用决策树法求解较为有效.

矩阵法(matrix method)　风险型决策问题中的一般方法. 设 $\theta=\{\theta_1,\theta_2,\cdots,\theta_m\}$为自然状态集，$\theta_i(i=1,2,\cdots,m)$为可能出现的重要的自然或环境的状态；$a=\{a_1,a_2,\cdots,a_n\}$为行动方案集，$a_j(j=1,2,\cdots,n)$为决策者可选择的一个行动；$P(\theta_i)=P_i$ 为状态 θ_i 发生的概率，称 $\boldsymbol{P}=(P(\theta_1),P(\theta_2),\cdots,P(\theta_m))$为其状态概率向量，且

$$\sum_{i=1}^{m} P_i = 1.$$

设风险矩阵或益损矩阵为 $A=(a_{ij})_{m\times n}$，则 a_j 的益损期望值为

$$E(a_j) = \sum_{i=1}^{m} P_i a_{ij},$$

且 $AP^{\mathrm{T}}=E(a)$. 若决策目标为收益最大，则

$$a_r = \max_{1\leqslant j\leqslant n}(E(a_j))$$

为最优方案；若决策目标为损失最小，则

$$a_s = \min_{1\leqslant j\leqslant n}(E(a_j))$$

为最优方案. 当 $E(a_j)$ 有两个以上的相同值时，则比较期望值与它的收益值的下界差，即

$$D(a_j) = E(a_j) - \min_{1\leqslant i\leqslant m}(a_{ij}).$$

若 $D(a_j)$ 不同，选取小的；若 $D(a_j)$ 相同，则任选一个均可. 若 $E(a_j)$ 相同，则比较损失值的上界与期望值之差，即

$$D(a_j) = \max_{1\leqslant i\leqslant m}(a_{ij}) - E(a).$$

不确定型决策问题(nondeterministic type decision problem)　一类决策问题. 是自然或环境状态都为待定的决策问题.

乐观准则(optimistic criterion)　亦称最大-最大准则. 特殊的决策准则之一. 持乐观决策准则的决策者对风险的态度总是乐观的，对客观情况不明而需要作决策时，总是争取获得最好结果的机会，而从大中之大来选择他的决策策略. 其决策步骤为：

1. 求出每个方案在各种自然状态下的最大效益值.
2. 求各最大效益值的最大值.
3. 最大值对应的行动方案即为所求.

最大-最大准则(max-max criterion)　即"乐观准则".

悲观准则(pessimistic criterion)　亦称最小-最大准则、瓦尔德准则、保守决策准则. 特殊的决策准则之一. 持悲观决策准则的决策者对风险的态度总是悲观的，对客观各事件发生的概率不明而需作决策时，非常谨慎，先估计各种最坏的可能结果，然后从中选择最好的作为决策策略. 其决策步骤为：

1. 求出每个方案在各种自然状态下的最小效益值.
2. 求各最小效益值的最大值.
3. 最大值对应的行动方案，即为所求.

若已给的是损失矩阵，则应采取最大-最小准则.

最小-最大准则(min-max criterion)　即"悲观准则".

瓦尔德准则(Wald criterion)　即"悲观准则".

乐观系数准则(optimistic coefficient criterion)　亦称赫威斯准则. 一种决策调整准则. 决策者对用最大-最大准则和最小-最大准则进行决策皆不满意，从而采取一个既不偏大、也不偏小的折衷，即取一个乐观系数 $\alpha(0\leqslant\alpha\leqslant1)$，并且计算

$$H_i = \alpha a_{i\max} + (1-\alpha)a_{i\min},$$

式中 $a_{i\max}$ 和 $a_{i\min}$ 分别表示第 i 个策略可能得到的最大效益值和最小效益值. 最后比较 H_i 值，选择 $\max\{H_i\}$ 作为所求.

赫威斯准则(Hurwicz criterion)　即"乐观系数准则".

等可能性准则(equally liability criterion)　亦称拉普拉斯准则. 信息不完全时的一种决策准则. 此准则是数学家拉普拉斯(Laplace, P. S.)于 19 世纪提出的. 当决策者在决策过程中不能确定这一事件容易发生，还是那一事件容易发生，只好认为各种事件发生的机会是相等的，即发生的概率是相等的. 先计算各策略的 S_i 效益期望值 $E(S_i)$，然后在所有这些期望值中选择最大的，即按 $\max\limits_i\{E(S_i)\}$ 来决定决策策略.

拉普拉斯准则(Laplace criterion)　即"等可能性准则".

后悔值准则(regret value criterion)　亦称沙万奇准则或遗憾准则. 一种决策调整准则. 决策者在制定决策之后，其结果未能达到满意的要求，产生后悔感或遗憾感. 后悔值(或遗憾值)就是由于决策者没有选用收益最大的策略而形成的机会损失值. 设事件为 t，各策略的效益为 $a_{it}(i=1,2,\cdots,n)$，则效益最大的策略为 $a_{lt}=\max\limits_i(a_{it})$，其机会损失值为

$$a'_{lt} = \{\max_i(a_{it}) - a_{it}\}.$$

再计算

$$\max_j a'_{ij} = \max_j\{\max_i(a_{it}) - a_{it}\},$$

最后按 $\min\limits_i \max\limits_j a'_{ij}$，即可得所求的决策策略.

沙万奇准则(Savage criterion)　即"后悔值准则".

效用(utility)　一种表示决策价值的数值. 在有风险的情况下，为使决策者所作的决策符合其价值观，达到预期的最优效果，必须对各种决策后果(包括决策者对后果的偏爱)予以量化，即赋以某种数值，称为效用. 一般用 1 表示最大效用，用 0 表示最小效用. 称决策者对一不确定事件可能冒的风险的态度的效用为基数效用. 称决策者无需承担任何风险的效用为序数效用. 效用这个概念，首先是由丹尼尔第一·伯努利(Bernoulli, Daniel Ⅰ)于 1738 年提出的. 他认为人们对其钱财的真实价值的考虑与他的钱财拥有量之间有对数关系，即

$$U=\log_a M(a>1, M>0),$$

式中 U 表示效用，M 表示货币. 这就是伯努利的货币效用函数.

理性行为公理(reason behaviour axiom)　亦称偏爱结构公理. 是指决策者对供选方案集或后果集中各元素间的偏爱关系应满足的某些性质. 设 $x_1, x_2, \cdots, x_n$ 表示决策者选择某一供选(行动)方案时，决策的 n 个可能后果，$X=\{x_1, x_2, \cdots, x_n\}$ 表示决策后果集(亦称决策报酬集). 若决策者认为后果 x_1 偏爱(优)于后果 x_2，则记为 $x_1 \succ x_2$；若决策者认为后果 x_1 无差异(等价)于后果 x_2，则记为 $x_1 \sim x_2$. 对这 n 个可能后果，理性行为公理是：

1. $x_1 \succ x_2, x_1 \sim x_2, x_2 \succ x_1$，其中只有一个成立.

2. $x_1 \sim x_2$ 当且仅当 $x_2 \sim x_1$.

3. 若 $x_1 \succ x_2, x_2 \succ x_3$，则 $x_1 \succ x_3$.

4. 若 $x_1 \succ x_2$，则对任意的 $x_3 \in X$ 和 $p \in (0,1)$，必有 $p(x_1)+(1-p)(x_3) \succ p(x_2)+(1-p)(x_3)$.

5. 若 $x_1 \succ x_2 \succ x_3$，则存在 $p, p_1, p_2 \in (0,1)$，使得 $p_1(x_1)+(1-p_1)(x_3) \succ p(x_2), p(x_2) \succ p_2(x_1)+(1-p_2)(x_3)$，式中 p, p_1, p_2 均表示概率.

1976 年，霍华德(Howard, R. A.)、马斯孙(Matheson, J. E.)和米勒(Miller, R. E.)等人又增添了一条抽奖的 δ 性质公理(参见“展望集”).

展望集(envision set)　所有展望的集合. 设 $X=\{x_1, x_2, \cdots, x_n\}$ 是(决策)后果集，p_i 是 x_i 出现的概率($i=1,2,\cdots,n$)，且

$$\sum_{i=1}^{n} p_i = 1,$$

$\boldsymbol{p}=(p_1, x_1; p_2, x_2; \cdots, p_n, x_n)$ 表示后果 x_1 以概率 p_1 出现，后果 x_2 以概率 p_2 出现，…，后果 x_n 以概率 p_n 出现. 称 $\boldsymbol{p}$ 为展望，称 $\Phi=\{\boldsymbol{p}\}$ 为展望集. 展望集 Φ 具有以下性质：

1. 在凸线性组合下 Φ 是闭的，即

$$\lambda \boldsymbol{p}^1 + (1-\lambda)\boldsymbol{p}^2 \in \Phi,$$
$$\boldsymbol{p}^1 \in \Phi,\ \boldsymbol{p}^2 \in \Phi,\ \lambda \in [0,1].$$

2. 所有退化的概率分布属于 Φ.

抽奖是展望的特例，抽奖的奖金即抽奖事件的后果. 一抽奖的所有奖金都增加一金额 δ，将使此抽奖的确定当量增加 δ. 这就是抽奖的 δ 性质. 确定当量是指：一是决策者得到一确定的后果 c_1；另一是决策者得到一抽奖的机会，以概率 p 得到后果 c_2，以概率 $1-p$ 得到后果 c_3，即 $(p, c_2; (1-p), c_3)$. 若决策者认为这两种情况是等价的，则确定的后果 c_1 称为抽奖 $(p, c_2; (1-p), c_3)$ 的确定当量.

效用函数(utility function)　一种表示决策价值的函数. 设 X 为后果集(报酬集)，$u(\cdot)$ 为 X 上的实值函数. 若对于任意的 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in X$，有：$\boldsymbol{x}_1 \succ \boldsymbol{x}_2$ 当且仅当 $u(\boldsymbol{x}_1) > u(\boldsymbol{x}_2)$；$\boldsymbol{x}_1 \sim \boldsymbol{x}_2$，当且仅当 $u(\boldsymbol{x}_1) = u(\boldsymbol{x}_2)$，则称 $u(\cdot)$ 为决策问题的效用函数.

效用函数构造法(structure method of utility function)　构造效用函数的方法，也是一种求解最优决策的方法. 是决策者在决策之前，根据自己的价值观建立效用函数，从而寻求最优策略的方法. 其步骤为：

1. 选择参考点. 在后果集 X 中任选两个元素 $\boldsymbol{x}_1, \boldsymbol{x}_2$，且使 $\boldsymbol{x}_2 \succ \boldsymbol{x}_1$，令 $u(\boldsymbol{x}_1)=0, u(\boldsymbol{x}_2)=1$. 如果在偏爱结构中 $\boldsymbol{x}_1$ 是最差的后果，$\boldsymbol{x}_2$ 是最好的后果，建立效用函数就更方便.

2. 确定其他点的效用. 如果对于 $\boldsymbol{x}_3 \in X$，且 $\boldsymbol{x}_2 \succ \boldsymbol{x}_3 \succ \boldsymbol{x}_1$，找到一个数 $p \in (0,1)$，使得 $\boldsymbol{x}_3 \sim p(\boldsymbol{x}_1)+(1-p)(\boldsymbol{x}_2)$，根据效用函数的定义，$\boldsymbol{x}_3$ 的效用为

$$u(\boldsymbol{x}_3) = pu(\boldsymbol{x}_1) + (1-p)u(\boldsymbol{x}_2) = 1-p.$$

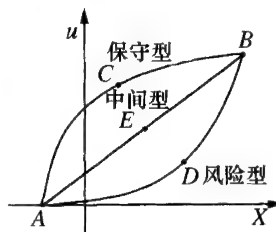
若对于 $\boldsymbol{x}_3 \in X$，且 $\boldsymbol{x}_1 \succ \boldsymbol{x}_3$，找到一个数 $p \in (0,1)$，使得 $\boldsymbol{x}_1 \sim p(\boldsymbol{x}_3)+(1-p)(\boldsymbol{x}_2)$，则由 $u(\boldsymbol{x}_1)=pu(\boldsymbol{x}_3)+(1-p)u(\boldsymbol{x}_2)=0$，即 $pu(\boldsymbol{x}_3)=-(1-p)u(\boldsymbol{x}_2)=p-1$，从而得 $u(\boldsymbol{x}_3)=(p-1)/p$. 若对于 $\boldsymbol{x}_3 \in X$，且 $\boldsymbol{x}_3 \succ \boldsymbol{x}_2$，找到一个数 $p \in (0,1)$，使得 $\boldsymbol{x}_2 \sim p(\boldsymbol{x}_1)+(1-p)(\boldsymbol{x}_3)$，则由 $u(\boldsymbol{x}_2)=pu(\boldsymbol{x}_1)+(1-p)u(\boldsymbol{x}_3)=(1-p)u(\boldsymbol{x}_3)=1$，解得

$$u(\boldsymbol{x}_3) = \frac{1}{(1-p)}.$$

3. 检查一致性. 用上述方法可以求得不同后果的效用，但必须检查已求效用的一致性. 假定已用上述方法求出了 $\boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5 \in X$ 的效用，且已知 $\boldsymbol{x}_5 \succ \boldsymbol{x}_4 \succ \boldsymbol{x}_3$，则存在 $q \in (0,1)$，使得 $u(\boldsymbol{x}_4)=qu(\boldsymbol{x}_3)+(1-q)u(\boldsymbol{x}_5)$，需要检查 $\boldsymbol{x}_4 \sim q(\boldsymbol{x}_3)+(1-q)(\boldsymbol{x}_5)$ 是否成立. 否则，需要重新求得效用以满足一致性要求.

4. 如果后果集 X 是连续点集，在求得适当个离散点的效用且满足一致性要求之后，利用平滑方法可以求得其余点的效用.

效用曲线(utility curve)　一种表示决策态度的曲线. 指在直角坐标系内，用决策后果集(报酬集)和效用表示决策者对风险态度变化关系的曲线. 一般用横坐标表示决策后果集 $X=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$，用纵坐标表示效用函数值 $u(\boldsymbol{x}_i)$. 效用曲线有三种类型，如图所示. 曲线 ACB 表示决策者厌恶风险，是一个不求大利、避免风险、谨慎小心的保守型决策者. 曲线 ADB 表示决策者追求风险，是一个谋求大利、不怕风险、富于进取的风险型决策者. 曲线 AEB 表示决策者对风险持中立态度，是一个中间型决策者.



效用曲线的拟合(fitting of utility curve)　将离散的效用信息用函数关系拟合成效用曲线. 有六种常用的拟合表达式：

1. 线性函数：$u(\boldsymbol{x})=c_1+a_1(\boldsymbol{x}-c_2)$.

2. 指数函数：$u(\boldsymbol{x})=c_1+a_1(1-e^{a_2(\boldsymbol{x}-c_2)})$.

3. 双指数函数：

$$u(\boldsymbol{x})=c_1+a_1(2-e^{a_2(\boldsymbol{x}-c_2)}-e^{a_3(\boldsymbol{x}-c_3)}).$$

4. 指数加线性函数：

$$u(\boldsymbol{x})=c_1+a_1(1-e^{a_2(\boldsymbol{x}-c_2)})+a_3(\boldsymbol{x}-c_2).$$

5. 幂函数：$u(\boldsymbol{x})=a_1+c_1a_2[c_1(\boldsymbol{x}-a_3)]^{a_4}$.

6. 对数函数：$u(\boldsymbol{x})=c_1+a_1\lg(c_3\boldsymbol{x}-c_2)$，

其中 $a_1,a_2,a_3,a_4,c_1,c_2,c_3$ 均为常数.

可测价值函数(measurable value function)　后果集 X 上的一个实值函数，用 v 表示. 它在 X 上可以起到序数效用函数的作用，但又和序数效用函数不同. 由于按照 v 的区间的偏爱程度之差的性质要求它对正线性变换不变，而不是对保序变换不变，因此称 v 为可测价值函数. 得叶(Dyer, I. S.)和萨林(Sarin, R. K.)曾在 1979 年和 1982 年分别研究了可测价值函数是在 X 上的一个四元关系，建立了这四元关系的公理系统，并证明了当优先关系 $\geqslant^*$ 满足这公理系统时，在 X 上存在一实值函数 v，对于所有的 $\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4\in X$，有 $\boldsymbol{x}_1\boldsymbol{x}_2\geqslant^*\boldsymbol{x}_3\boldsymbol{x}_4$ 当且仅当

$$v(\boldsymbol{x}_1)-v(\boldsymbol{x}_2)\geqslant v(\boldsymbol{x}_3)-v(\boldsymbol{x}_4),\ \boldsymbol{x}_1\boldsymbol{x}_2\geqslant^*\boldsymbol{x}_3\boldsymbol{x}_4$$

表示在 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 之间的偏爱程度之差超过了在 $\boldsymbol{x}_3$ 和 $\boldsymbol{x}_4$ 之间的偏爱程度之差. 此外，v 经过一正线性变换 $\theta(\boldsymbol{x})=\alpha v(\boldsymbol{x})+\beta(\alpha>0)$，对于所有的 $\boldsymbol{x}\in X$，优先关系 $\geqslant^*$ 是不变的.

风险态度的局部测度(local measure of risk manner)　决策者对待风险态度的一种度量. 设 $u(\boldsymbol{x})$ 是效用函数，它在后果集 X 上是单调递增的，且具有连续的二阶导数，$\boldsymbol{x}\in X$，则称

$$w(\boldsymbol{x})=-\frac{u''(\boldsymbol{x})}{u'(\boldsymbol{x})}$$

为风险态度的局部测度. 若 $w(\boldsymbol{x})>0$，则决策者是厌恶风险的；若 $w(\boldsymbol{x})<0$，则决策者是追求风险的；若 $w(\boldsymbol{x})=0$，则决策者对风险持中立态度.

偏爱程度的局部测度(local measure of preference degree)　决策者偏爱程度的一种度量. 设 v 是可测价值函数，它在后果集 X 上是单调增加的，且具有连续的二阶导数，$\boldsymbol{x}\in X$，则称

$$m(\boldsymbol{x})=-\frac{v''(\boldsymbol{x})}{v'(\boldsymbol{x})}$$

为偏爱程度的局部测度. 当 $m(\boldsymbol{x})>0$ 时，在 $\boldsymbol{x}$ 有递减的边际价值；当 $m(\boldsymbol{x})<0$ 时，在 $\boldsymbol{x}$ 有递增的边际价值；当 $m(\boldsymbol{x})=0$ 时，在 $\boldsymbol{x}$ 有不变的边际价值.

相对风险态度的局部测度(local measure of relative risk manner)　风险态度的局部测度和偏爱程度的局部测度之间大小关系的一种度量. 设 u 和 v 定义在 $X'\subset X$ 内，令 $\boldsymbol{x}\in X'$(参见“风险态度的局部测度”和“偏爱程度的局部测度”). 若 $m(\boldsymbol{x})>w(\boldsymbol{x})$，则决策者是相对追求风险的；若 $m(\boldsymbol{x})<w(\boldsymbol{x})$，则决策者是相对厌恶风险的；若 $m(\boldsymbol{x})=w(\boldsymbol{x})$，则决策者是相对风险中立的. 若对于所有的 $\boldsymbol{x}\in X'$，$w(\boldsymbol{x})>m(\boldsymbol{x})>0$，则 u 和 v 都是凹的，但 u 比 v 更凹. 反之，若 $w(\boldsymbol{x})<m(\boldsymbol{x})<0$，则 u 和 v 都是凸的，但 u 比 v 更凸. 令

$$u_v[v(\boldsymbol{x})]=u(\boldsymbol{x}),\ w_v[v(\boldsymbol{x})]=-\frac{u''_v[v(\boldsymbol{x})]}{u'_v[v(\boldsymbol{x})]},$$

式中微分是对 v 取的，则有下面的结论：在 $\boldsymbol{x}$ 上，决策者是相对追求风险的当且仅当 $w_v[v(\boldsymbol{x})]<0$；决策者是相对厌恶风险的当且仅当 $w_v[v(\boldsymbol{x})]>0$；决策者是相对风险中立的当且仅当 $w_v[v(\boldsymbol{x})]=0$.

损失函数(loss function)　一种决策后效函数. 指一个损失函数 $L(\theta,a)$ 表示当状态为 θ、决策者的行动为 a 时，所产生的后果使决策者遭受的损失. 因后果的效用愈大，损失愈小，故可用效用函数 u 定义损失函数 L，即损失函数定义为

$$L(\theta',a')=\sup_{\theta\in\Theta}\sup_{a\in A}u(\theta,a)-u(\theta,a),$$

式中 Θ 表示状态集，A 表示行动集.

风险函数(risk function)　损失函数对后果的期望值. 设行动 $a\in A$ 是后果 $\boldsymbol{x}\in X$ 的函数，记为 $\gamma(\boldsymbol{x})$，则有

$$R(\theta,\gamma(\boldsymbol{x}))=E_\theta^{\boldsymbol{x}}L(\theta,\gamma(\boldsymbol{x})),$$

式中 E 的上标 $\boldsymbol{x}$ 表示对 $\boldsymbol{x}$ 的数学期望，下标 θ 表示对 θ 的数学期望. R 称为 $\gamma(\cdot)$ 的风险函数，又称为决策律的风险. 若 $\boldsymbol{x}$ 为连续随机变量，则

$$R(\theta,\gamma(\boldsymbol{x}))=\int_{\boldsymbol{x}\in X}L(\theta,\gamma(\boldsymbol{x}))f(\boldsymbol{x}|\theta)\mathrm{d}\boldsymbol{x};$$

若 x 为离散随机变量，则

$$R(\theta,\gamma(\boldsymbol{x}))=\sum_{\boldsymbol{x}\in X}L(\theta,\gamma(\boldsymbol{x}))f(\boldsymbol{x}|\theta).$$

贝叶斯风险(Bayes risk)　一种含先验信息的决策风险. 风险函数 $R(\theta^*,\gamma(\boldsymbol{x}))$ 对 θ^* 取期望值，记为 $r(\pi,\gamma(\boldsymbol{x}))$，即

$$\begin{aligned}r(\pi,\gamma(\boldsymbol{x}))&=E^\pi R(\theta^*,\gamma(\boldsymbol{x}))\\&=E^\pi(E_{\theta^*}^\pi L(\theta^*,\gamma(\boldsymbol{x}))),\end{aligned}$$

式中 θ^* 表示随机状态，$\pi(\theta)$ 是随机状态 θ^* 的先验密度. 称 $r(\pi,\gamma(\boldsymbol{x}))$ 为贝叶斯风险. 若 θ^* 为连续的随机变量，则

$$\begin{aligned}&r(\pi,\gamma(\boldsymbol{x}))\\&=\int_{\theta\in\Theta}R(\theta,\gamma(\boldsymbol{x}))\pi(\theta)\mathrm{d}\theta\\&=\int_{\theta\in\Theta}\int_{\boldsymbol{x}\in X}L(\theta,\gamma(\boldsymbol{x}))f(\boldsymbol{x}|\theta)\pi(\theta)\mathrm{d}\boldsymbol{x}\mathrm{d}\theta;\end{aligned}$$

若 θ^* 为离散的随机变量，则

$$\begin{aligned}r(\pi,\gamma(\boldsymbol{x}))&=\sum_{\theta\in\Theta}R(\theta,\gamma(\boldsymbol{x}))\pi(\theta)\\&=\sum_{\theta\in\Theta}\sum_{\boldsymbol{x}\in X}L(\theta,\gamma(\boldsymbol{x}))f(\boldsymbol{x}|\theta)\pi(\theta).\end{aligned}$$

贝叶斯原理(Bayes principle)　亦称贝叶斯决策原则.一种决策原则.是用贝叶斯风险表示的最优决策律.若存在一决策律 γ_1,使得对一切 γ_2 均有 $r(\pi,\gamma_1)\leqslant r(\pi,\gamma_2)$,则称 γ_1 是用贝叶斯风险表示的最优决策律,即贝叶斯原理.

贝叶斯规则(Bayes rule)　一类最优决策规则.此规则是:按照贝叶斯原理,若在决策规则集 Γ 中的决策规则的偏爱(优先)关系满足理性行为公理(参见"理性行为公理"),则在 Γ 中必存在一最优决策律 γ^*,使 $r(\pi,\gamma)$ 达到极小,即有

$$r(\pi,\gamma^*)=\inf_{\gamma\in\Gamma}r(\pi,\gamma).$$

决策分析(decision analysis)　一种学科名称或一种决策手段.作为学科名称,它指决策论中研究含有不确定性的决策问题的理论和方法部分(参见"决策论").有时,(狭义意义下的)决策论即指决策分析.作为决策的手段,它指用决策原理去分析决策问题.决策分析分为两类:一类是利用状态的先验信息,而不利用抽样试验结果,称此为先验分析;另一类是利用抽样试验的结果得到状态的后验概率,再使后验期望损失为最小,称此为后验分析.用先验分析和后验分析所得的结果,一般是不相同的.决策分析有参数估计型决策分析和假设检验型决策分析两种.

参数估计型决策分析(decision analysis with parameter estimation)　一类统计决策分析方法.指对未知参数 θ 的估计,这个估计与损失函数的最优估计有着密切的关系.常用的最优估计有下面几种:

1.平方损失的最优估计.设后验概率密度为 $\pi(\theta|x)$,则后验损失为

$$\begin{aligned}E^{\pi(\theta|x)}[L(\theta,a)]&=E^{\pi(\theta|x)}[(\theta-a)^2]\\&=[a-E^{\pi(\theta|x)}(\theta)]^2+E^{\pi(\theta|x)}(\theta^2)\\&\quad-[E^{\pi(\theta|x)}(\theta)]^2.\end{aligned}$$

要使该式达到最小当且仅当 $a^*=\gamma(\boldsymbol{x})=E^{\pi(\theta|\boldsymbol{x})}(\boldsymbol{x})$.此即平方损失条件下的最优决策.

2.加权平方损失的最优估计.即

$$\begin{aligned}&E^{\pi(\theta|x)}[L(\theta,a)]\\&=E^{\pi(\theta|x)}[w(\theta)(\theta-a)^2]\\&=\left[a-\frac{E^{\pi(\theta|x)}[w(\theta)\theta]}{E^{\pi(\theta|x)}[w(\theta)]}\right]^2E^{\pi(\theta|x)}[w(\theta)]\\&\quad+E^{\pi(\theta|x)}[w(\theta)\theta^2]-\frac{(E^{\pi(\theta|x)}[w(\theta)\theta])^2}{E^{\pi(\theta|x)}[w(\theta)]},\end{aligned}$$

相应的最优决策为

$$a^*=\gamma(\boldsymbol{x})=\frac{E^{\pi(\theta|x)}[w(\theta)\theta]}{E^{\pi(\theta|x)}[w(\theta)]}.$$

3.线性损失的最优估计.设最优决策是 a^*,同时假定 A 中的策略可以比较大小,则

$$\begin{aligned}&L(\theta,a^*)-L(\theta,a)\\&=\begin{cases}k_1(a^*-a)&(\theta\leqslant a^*),\\(k_0+k_1)\theta-(k_0a^*+k_1a)&(a^*<\theta\leqslant a),\\k_0(a-a^*)&(a<\theta),\end{cases}\end{aligned}$$

且

$$\begin{aligned}L(\theta,a^*)-L(\theta,a)\leqslant&k_1(a^*-a)I_{(-\infty,a^*)}(\theta)\\&+k_0(a-a^*)I_{(a^*,\infty)}(\theta),\end{aligned}$$

式中 $I_{(a,b)}(\theta)$ 表示 I 在 $\theta\in(a,b)$ 上的值为1,在实轴上除此之外部分的值为0.若 a^* 满足

$$P(\theta\leqslant a^*|\boldsymbol{x})=\frac{k_0}{k_1+k_0},$$

$$P(\theta>a^*|\boldsymbol{x})=\frac{k_1}{k_1+k_0},$$

则有

$$E^{\pi(\theta|x)}[L(\theta,a^*)]\leqslant E^{\pi(\theta|x)}[L(\theta,x)],$$

式中 a^* 是概率密度 $\pi(\theta|\boldsymbol{x})$ 的 $k_0/(k_0+k_1)$ 分位数.对于绝对值损失,只要 $k_0=k_1$,其最优决策为 $a^*=\pi(\theta|\boldsymbol{x})$ 的1/2分位数.

假设检验型决策分析(decision analysis with hypothesis testing)　一种常见的统计决策分析.设 Θ_0,Θ_1 分别表示不同的自然状态集;假设 H_0 表示 $\theta\in\Theta_0$,H_1 表示 $\theta\in\Theta_1$;策略空间 $A=\{a_0,a_1\}$ 有两个元素,a_0 表示接受假设 H_0,a_1 表示接受假设 H_1.一般假定 Θ_0,Θ_1 均为自然状态空间 Θ 的子集,且 $\Theta_0\cup\Theta_1=\Theta,\Theta_0\cap\Theta_1=\varnothing$.决策损失函数是所谓"0-$k_i$"型损失,即

$$L(\theta,a_i)=\begin{cases}0,&\theta\in\Theta_i,\\k_i,&\theta\in\Theta_i^c\quad(i=0,1),\end{cases}$$

式中 $\Theta_i=\Theta_0$(或 Θ_1),则 $\Theta_i^c=\Theta_1$(或 Θ_0).若 θ 的后验概率是 $\pi(\theta|\boldsymbol{x})$,后验期望损失为

$$E^{\pi(\theta|x)}L(\theta,a_i)=k_iP(\Theta_i^c|\boldsymbol{x})(i=0,1).$$

若 $k_0P(\Theta_1|\boldsymbol{x})>k_1P(\Theta_0|\boldsymbol{x})$ 或

$$\frac{k_0}{k_1}>\frac{P(\Theta_0|\boldsymbol{x})}{P(\Theta_1|\boldsymbol{x})},$$

则接受 H_1;若 $k_0P(\Theta_1|\boldsymbol{x})<k_1P(\Theta_0|\boldsymbol{x})$ 或

$$\frac{k_0}{k_1}<\frac{P(\Theta_0|\boldsymbol{x})}{P(\Theta_1|\boldsymbol{x})},$$

则接受 H_0.根据对立事件的概率公式 $P(\Theta_0|\boldsymbol{x})=1-P(\Theta_1|\boldsymbol{x})$,则上面两个结论分别变为:若

$$P(\Theta_1|\boldsymbol{x})>\frac{k_1}{k_0+k_1}\left(\text{或 }P(\Theta_0|\boldsymbol{x})<\frac{k_0}{k_1+k_0}\right),$$

则接受假设 H_1;若

$$P(\Theta_1|\boldsymbol{x})<\frac{k_1}{k_0+k_1}\left(\text{或 }P(\Theta_0|\boldsymbol{x})>\frac{k_0}{k_1+k_0}\right),$$

则接受假设 H_0.

极小化极大原则(minimax principle)　一类决策原则.设风险函数有关系

$$\sup_{\theta\in\Theta}R(\theta,\gamma_1^*)<\sup_{\theta\in\Theta}R(\theta,\gamma_2^*),$$

式中

$$\sup_{\theta\in\Theta} R(\theta,\gamma^*)$$

表示采用决策规则 γ^* 时遇到最坏的风险，则认为随机决策规则 γ_1^* 优于随机决策规则 γ_2^*. 从随机决策规则集 Γ^* 中选择一个决策规则 γ^{*m}，使

$$\sup_{\theta\in\Theta} R(\theta,\gamma^{*m})$$

为极小，即

$$\sup_{\theta\in\Theta} R(\theta,\gamma^{*m}) = \inf_{\gamma^*\in\Gamma^*}\sup_{\theta\in\Theta} R(\theta,\gamma^*),$$

称此决策规则为极小化极大原则. 有的问题可用极小化极大原则去确定一最优非随机决策规则. 若这个最优决策规则存在，称它为极小化极大非随机决策规则.

允许决策规则(permissible decision rule)　判定决策的一种方法. 在决策规则集 Γ 中没有其他决策规则优于决策规则 γ，则称 γ 为允许决策规则. 否则，称 γ 为不允许决策规则.

贝叶斯分析的扩展型(expand type of Bayes analysis)　一种决策模型. 指由式

$$(m(\boldsymbol{x}))^{-1}\int_\Theta L(\theta,a)f(\boldsymbol{x}|\theta)\pi(\theta)\mathrm{d}\theta$$
$$=\int_\Theta L(\theta,a)\pi(\theta|\boldsymbol{x})\mathrm{d}\theta$$

去寻求形式贝叶斯规则的模型，式中 $m(\boldsymbol{x})$ 是预测分布，$\pi(\theta|\boldsymbol{x})$ 是后验密度. 若相对于每个给定的 $\boldsymbol{x}$ 选择一行动 a，使

$$\int_\Theta L(\theta,a)f(\boldsymbol{x}|\theta)\pi(\theta)\mathrm{d}\theta$$

为极小，它必使后验期望损失

$$\int_\Theta L(\theta,a)\pi(\theta|\boldsymbol{x})\mathrm{d}\theta$$

为极小. 由此有：若对于每个 $\boldsymbol{x}$ 选择一行动 a，使后验期望损失为极小，或等价地使

$$\int_\Theta L(\theta,a)f(\boldsymbol{x}|\theta)\pi(\theta)\mathrm{d}\theta$$

为极小，就能找到贝叶斯行动 a 和对应的贝叶斯规则 $\gamma(\boldsymbol{x})$，称此规则为形式贝叶斯规则，相应的行动称为贝叶斯行动.

先验信息(prior information)　主观因素的概率. 自然状态 θ 的先验信息，其表现形式是 θ 的先验概率，是进行贝叶斯决策的基本要素之一. 在一些决策问题中，θ 具有随机性；在另一些决策问题中，θ 不具有明显的随机性，这时，其先验概率由决策者对 θ 值出现各种可能性的主观可信程度来确定，把这种先验概率称为"主观概率".

确定先验概率的方法(method of determination prior probability)　确定有信息先验概率的一些方法. 有信息是指决策者已积累了处理类似决策问题的经验，或者通过有关专家咨询获得了对自然状态 θ 的某些看法. 有下面四种方法：

1. 比较法. 先比较自然状态 θ_1 和 θ_2 出现的先验概率 $\pi(\theta_1)$ 和 $\pi(\theta_2)$ 的大小，若

$$\pi(\theta_1)\geqslant\frac{1}{2},\quad \pi(\theta_2)<\frac{1}{2},$$

再分析 θ_1 出现的可能性是否大于 θ_2 出现可能性的 2 倍，若是，则有

$$\pi(\theta_1)\geqslant\frac{2}{3},\quad \pi(\theta_2)<\frac{1}{3}.$$

直到把 $\pi(\theta_1)$ 取值的范围缩小到合理的程度. 这时，便求出了自然状态 θ 的先验概率. 设

$$\pi(\theta_1)=p,\ \pi(\theta_2)=1-p\quad(0\leqslant p\leqslant 1),$$

同时，对由此得到的概率必须按 θ_2 进行检查. 若 $\pi(\theta_2)$ 的值不合理，还须调整，确保其一致性. 若自然状态 θ 的取值多于两个不同的有限值，即 $\theta=\{\theta_1,\theta_2,\cdots,\theta_n\}$，此时通过比较法确定 $\pi(\theta_i)=p_i(i=1,2,\cdots,n)$，但须注意满足

$$\sum_{i=1}^n p_i=1(p_i\geqslant 0).$$

当 n 比较大时，应采用方法 2.

2. 分解-综合比较法. 若 $\theta_i(i=1,2,\cdots,n)$ 均可视为两种或两种以上独立参量 $\omega_j(j=1,2,\cdots,m)$，$\upsilon_k(k=1,2,\cdots,s)$ 等的组合，即 $\theta_i=(\omega_j,\upsilon_k)$，且 m,s 均小于 n，则先用比较法确定 $\pi(\omega_j)\pi(\upsilon_k)$，再根据独立随机变量交集的公式有 $\pi(\theta_i)=\pi(\omega_j)\pi(\upsilon_k)\cdots$.

3. 分位法. 设 m_k 是一个实数，且有 $F(m_k)=P(\theta:\theta\leqslant m_k)=k(0\leqslant k\leqslant 1)$，则称 m_k 为 θ 的概率密度函数 $\pi(\theta)$ 的 k 分位数，式中 F 表示 θ 的分布函数. 称 $\pi(\theta)$ 的 1/2 分位数 $m_{\frac{1}{2}}$ 为中位数，$m_{\frac{1}{4}}$ 为下 1/4 分位数，$m_{\frac{3}{4}}$ 为上 1/4 分位数等. 若 $\theta=[a,b]\subset R$，则有

$$P\left(\theta:a\leqslant\theta\leqslant m_{\frac{1}{2}}\right)=P\left(\theta:m_{\frac{1}{2}}<\theta\leqslant b\right)=\frac{1}{2},$$
$$P\left(\theta:a<\theta\leqslant m_{\frac{1}{4}}\right)=P\left(\theta:m_{\frac{1}{4}}<\theta\leqslant m_{\frac{1}{2}}\right)$$
$$=P\left(\theta:m_{\frac{1}{2}}<\theta\leqslant m_{\frac{3}{4}}\right)$$
$$=P\left(\theta:m_{\frac{3}{4}}<\theta\leqslant b\right)=\frac{1}{4}$$

……………………

决策者可根据先验信息确定中位数和上、下 1/4 分位数等，再根据 θ 分布函数的光滑性、严格递增性等基本性质，求先验分布 $F(\theta)$. 据此，可求出概率密度函数 $\pi(\theta)$.

4. 曲线拟合法. 分析先验分布属于那种已知的分布函数族，由已经得到的分位数近似地确定参数，再确定所求的先验分布是该分布族中的某条曲线.

预后验分析(prediction posterior analysis)　亦称试验设计. 一种先行的统计分析. 为获得自然状态 θ 的更多信息，需要进行抽样试验. 为判断是否有

必要进行抽样试验以及决定试验次数，必须进行试验设计，即预后验分析.

固定试验次数法(method of fix experiment number)　亦称固定样本容量法. 简称固定次数法. 一种试验设计方法：在进行试验设计(即预后验分析)时，预先规定抽样试验次数为 n，相应的试验结果为 $x^n=(x_1,x_2,\cdots,x_n)$，全部做完这 n 次试验后再进行贝叶斯决策分析. 设 $\pi(\theta)$ 是 θ 的先验概率密度，$\gamma_n(\cdot)$ 是相对 n 次试验的贝叶斯决策律，则贝叶斯风险为

$$r^n(\pi)=E^{\pi(\theta)}E^{x^n}[L(\theta,\gamma_n(x^n),n)]$$
$$=E^{\pi(\theta)}E^{x^n}[L(\theta,\gamma_n(x^n))+K(n)].$$

设使上式取最小值的 n 为最优试验次数，则有

$$r^n(\pi)=E^{m(x^n)}E^{\pi(\theta|x^n)}[L(\theta,\gamma_n(x^n))+K(n)],$$

式中 $m(x^n)$ 表示 x^n 的边际密度，$\pi(\theta|x^n)$ 表示 θ 的条件密度，$L(\theta,\gamma_n(x^n))$ 表示预后验分析中的总损失，$K(n)$ 表示试验费用.

序贯试验法(sequential experiment method)　一种试验设计方法. 不预先规定试验次数，但每做一次或几次试验后，则按停止规则判别是否继续试验，直到符合停止规则为止，然后进行决策分析. 在符合停止规则之前，由于试验持续进行，从而分析也可以持续进行，称此分析为序贯分析.

完全信息期望值(expected value of completely information)　决策论的基本概念之一. 在没有完全信息时的期望损失与在理想状态下的期望损失之差. 简记为 EVCI：

$$\mathrm{EVCI}=\min_{a\in A}E^{\pi(\theta)}[L(\theta,a)]-E^{\pi(\theta)}[\min_{a\in A}L(\theta,a)].$$

抽样信息期望值(expected value of sample information)　简记为 EVSI. 决策论的基本概念之一. 极小贝叶斯风险和极小期望值之差，即

$$\mathrm{EVSI}=\min_{a\in A}E^{\pi(\theta)}[L(\theta,a)]$$
$$-\min_{\gamma\in\Gamma}E^{\pi(\theta)}[E_\theta^x L(\theta,\gamma(x))].$$

序贯决策步骤(sequential decision step)　简称序贯步骤. 一种决策分析方法. 其决策步骤是：在未试验前，采取一行动的贝叶斯风险比在进行试验(包括多次和无穷多次)后采取行动的贝叶斯风险要小，则不继续作试验，并立即采取行动；否则，再作一次试验，根据试验结果再作判断；这个步骤继续下去，直到试验可以停止为止，记为 $d(\tau,\gamma)$. 其中的 τ 表示停止规则，包含有函数 $\tau_0,\tau_1(\boldsymbol{x}),\tau_2(\boldsymbol{x}^2),\cdots,\tau_i(\boldsymbol{x}^i)$. $\tau_i(\boldsymbol{x}^i)$ 表示进行了 i 次观察，其实际观察值为 $\boldsymbol{x}^i=(x_1,x_2,\cdots,x_i)$ 时停止继续观察的概率. 在非随机情况，它们是 0 或 1，0 表示继续观察，1 表示停止观察. 在随机情况，它们是 0 与 1 之间的某个值，这个值表示停止观察的概率. 式中的 γ 为决策规则，包含有 $\gamma_0,\gamma_1(\boldsymbol{x}^1),\gamma_2(\boldsymbol{x}^2),\cdots,\gamma_i(\boldsymbol{x}^i)$. $\gamma_i(\boldsymbol{x}^i)$ 表示停止观察以后，根据实际观察 $\boldsymbol{x}^i$ 采取的行动. 序贯步骤在有限步停止的充分必要条件为：$P_\theta(N<+\infty)=1$ $(\forall\theta\in\Theta)$，而

$$P_\theta(N<+\infty)=P_\theta(N=0)+\sum_{n=1}^{\infty}P_\theta(N=n)$$
$$=P_\theta(N=0)+\sum_{n=1}^{\infty}\int_{\{N=n\}}\mathrm{d}F_n(\boldsymbol{x}^n|\theta),$$

式中的 $P_\theta(N=n)$ 是状态为 θ 时，决策过程在第 n 步停止的概率，$F_n(\boldsymbol{x}^n|\theta)$ 为给定 θ 时在集 X^n 上的条件分布函数. 序贯步骤 d 的风险函数为期望损失

$$L(\theta,d)=P(N=0)L(\theta,\gamma_0,0)$$
$$+\sum_{n=1}^{\infty}\int_{\{N=n\}}L(\theta,\gamma_n(\boldsymbol{x}^n),n)\mathrm{d}F_n(\boldsymbol{x}^n|\theta),$$

式中的 $\gamma_n(\boldsymbol{x}^n)$ 为前 n 步的实际观察值为 $\boldsymbol{x}^n$ 时的决策规则.

贝叶斯序贯步骤(sequential step of Bayes)　一种含先验信息的决策分析方法. 是使序贯步骤 d 的贝叶斯风险 $r(\pi,d)=E^{\pi(\theta)}[r(\theta,d)]$ 达到极小的方法，记为 $d^{\pi(\theta)}=(\tau^{\pi(\theta)},\gamma^{\pi(\theta)})$. 和 $d^{\pi(\theta)}$ 对应的贝叶斯风险为

$$r(\pi(\theta))=\inf_d r(\pi(\theta),d).$$

称

$$r_0(\pi(\theta)^n,n)=\inf_{a\in A}r_0(\pi(\theta)^n,a,n)$$
$$=\inf_{a\in A}E^{\pi(\theta)^n}[L(\theta,a,n)]$$

为后验贝叶斯风险，式中的 $\pi(\theta)^n$ 表示先验密度为 $\pi(\theta)$，实际观察值为 x^n 的后验密度. 若在 n 次观察后再继续观察 $x_{n+1},x_{n+2},\cdots$，并把前 n 次观察后得到的后验密度 $\pi(\theta)^n$ 作为第 $n+1$ 步开始的先验密度，则期望的贝叶斯风险记作 $r(\pi(\theta)^n,d,n)$. 若用 D^n 记第 $n+1$ 步开始后所有的序贯步骤的集，则极小的期望贝叶斯风险为

$$r(\pi(\theta)^n,n)=\inf_{d\in D^n}r(\pi(\theta)^n,d,n).$$

截尾步骤(curtailed step)　一种决策分析方法. 指抽样最多允许 m 步(包括已观察的 n 次)的方法为 m 截尾步骤. 用 D_m^n 记 m 截尾步骤 d 的集，称

$$r_{m-n}(\pi^n,n)=\inf_{d\in D_m^n}r(\pi^n,d,n)$$

为从第 n 步开始的序贯试验的 m 截尾贝叶斯风险. m 截尾的停止规则为：在第 n 步，计算其后验贝叶斯风险 $r_0(\pi^n,n)$，并把它和从第 n 步开始的序贯试验的 m 截尾贝叶斯风险 $r_{m-n}(\pi^n,n)$ 作比较. 若

$$r_0(\pi^n,n)=r_{m-n}(\pi^n,n),$$

则观察停止，并立即采取一行动，否则继续观察一次. 设贝叶斯规则 γ_n^π 对所有的 n 存在，而且函数 $r_j(\pi^n,n)$ 对于所有 $j\leqslant m$ 和 $n\leqslant m-j$ 是有限的，则一

贝叶斯 m 截尾步骤 d^m 被给定为 $d^m=(\tau^m,\gamma^m)$，其中 γ 为决策规则，且 $\gamma^m=(\gamma_{n+1}^{\pi},\gamma_{n+2}^{\pi},\cdots,\gamma_m^{\pi})$，而 τ^m 为停止规则.并规定：若

$$r_0(\pi^n,n)=r_{m-n}(\pi^n,n)\quad(n=0,1,2,\cdots,m),$$

则进行前 n 次观察后停止观察，并采取行动.这个停止规则的计算步骤是：在第 0 步，把贝叶斯风险 $r_0(\pi,0)$ 和 m 截尾的贝叶斯风险 $r_m(\pi,0)$ 作比较，若两者相等，则停止观察并作一决定.若 $r_0(\pi,0)$ 大于 $r_m(\pi,0)$，则继续观察一次，由观察值 x 去计算后验贝叶斯风险 $r_0(\pi^1,1)$，并把它和 m 截尾贝叶斯风险 $r_{m-1}(\pi^1,1)$ 作比较，若它们相等，则停止观察并采取行动，否则，再继续观察一次.运用 m 截尾步骤时，还需弄清如何计算 m 截尾贝叶斯风险 $r_j(\pi^n,n)$.设 $r_j(\pi^n,n)$ 是有限的，它们能通过关系

$$r_j(\pi^n,n)=\min\{r_{j-1}(\pi^n,n),\ E^*[r_{j-1}(\pi^n(\theta|x_{n+1}),n+1)]\}$$

归纳地去计算，式中 $\pi^n(\theta|x_{n+1})$ 是观察 x_{n+1} 后的后验，而 E^* 是已知 x^n 时对 x_{n+1} 的预测分布期望值.该预测分布的密度为

$$m^*(x_{n+1})=\frac{m_{n+1}(x^{n+1})}{m_n(x^n)}\quad(m_n(x^n)>0).$$

有价证券问题的模型(model of negotiable securities problem) 亦称麦克维支模型.一类特殊决策模型.求一组变量 x_i 的值，使目标函数

$$E=\sum_{i=1}^{n}E_ix_i$$

的值达到极大，且满足约束条件

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}x_ix_j=V,\quad \sum_{i=1}^{n}x_i=1.$$

$(x_1,x_2,\cdots,x_n)$ 还受到其他约束，式中的 E_i 是 R_i 的期望值，σ_{ij} 是 R_i 和 R_j 的协方差，E 是 R 的期望值，即有价证券的平均回收，V 是 R 的方差.称 $(x_1,x_2,\cdots,x_n)$ 为有价证券混合，它的集被约束于

$$\sum_{i=1}^{n}x_i=1,$$

x_i 表示资金分配到产业 i 的百分比.有价证券的投资回收为

$$R=\sum_{i=1}^{n}R_ix_i,$$

R_i 表示决策者在固定的投资周期内投资于产业 i 的估计的每元收益，n 表示决策者投资于 n 种产业.1953 年，数理经济学家麦克维支(Markowitz, H.)提出了上述模型.

方差排序(variance arrange order) 亦称 EV 法或 EV 规则.求解有价证券问题的方法.其中 E 表示投资者对有价证券问题所具有的效用，是有价证券收益的平均值；V 表示有价证券收益的方差.EV 法的基本思想是：一项投资既希望它的期望收益最大，又希望它所遇到的风险最小.但上述二者往往难以兼顾.设以事件的概率分布的平均值 μ 表示某一随机事件的收益，风险正比于分布的均方差 σ，厌恶风险的投资者在收益的两种分布函数 F 和 G 之间选择，F 的平均值至少和 G 相等，F 的均方差不大于 G 的均方差，即设 $\mu_F\geqslant\mu_G,\sigma_F\leqslant\sigma_G$，则 F 优于 G，至少 F 不会劣于 G.若 $\mu_F>\mu_G,\sigma_F<\sigma_G$，则 F 严格优于 G，而 F 称为按 EV 的意义较 G 有优势，G 即使被取消也不会使投资者受到损失.若两个不等关系中只有一个成立，则行动的选择将取决于在平均值和均方差之间的权衡，按 EV 优势的判据，F 和 G 都不能取消.

随机优势(stochastic dominance) 依期望效用大小对分布函数优劣的衡量.指对于所有的 $u\in U_i(i=1,2,3)$，$F>_iG$ 当且仅当 $E(u,F)>E(u,G)$，式中 F 和 G 分别表示随机事件 X 和 Y 的分布函数，$E(u,F)$ 和 $E(u,G)$ 分别表示具有分布 F 和 G 的随机事件 X 和 Y 的期望效用.称 $F>_iG$ 为分布 F 对 G 的第一、第二和第三等的随机优势.随机优势具有非对称性、传递性和 $F>_1G$ 包含有 $F>_2G$，而 $F>_2G$ 又包含有 $F>_3G$ 这三条性质.设

$$F^{(1)}(x)=F(x),$$

$$F^{(2)}(x)=\int_0^x F^{(1)}(y)\mathrm{d}y,$$

$$F^{(3)}(x)=\int_0^x F^{(2)}(y)\mathrm{d}y,$$

当 $x\in I$，而 $I=[0,+\infty)$ 或 $I=[0,1]$，则有 $F>_1G$ 当且仅当 $G(x)\geqslant F(x)(\forall x\in I)$；$F>_2G$ 当且仅当 $G^{(2)}(x)\geqslant F^{(2)}(x)(\forall x\in I)$；$F>_3G$ 当且仅当 $\mu_F\geqslant\mu_G,G^{(3)}(x)\geqslant F^{(3)}(x)(\forall x\in I)$.随机优势排序和 EV 排序有以下关系：

1. 若 $\mu_F=\mu_G$，则 $F>_2G$ 或 $F>_3G$ 的必要条件是 $\sigma_F\leqslant\sigma_G$；在 $>_2$ 的情况下，应有严格的不等关系.

2. 若 $\mu_F>\mu_G$，则 $\sigma_F\leqslant\sigma_G$ 对于 $F>_1G$ 或 $F>_2G$，或 $F>_3G$，既不必要也不充分.

3. 对于 F 和 G，若 σ_F 和 σ_G 是均方差的单调递增函数，则 EV 规则等价于第二等随机优势.对正态分布的随机变量，EV 规则和第二等随机优势相同.

随机决策分析的步骤(step of stochastic decision analysis) 对含随机变量的决策问题进行决策分析的几个阶段.分为三个阶段，即确定阶段、随机阶段和信息阶段：

1. 在确定阶段中，不仅要建立模型，而且要对模型进行分析.分析观察变量的变化对后果值的影响，即作灵敏度分析.对决策变量进行灵敏度分析，就是把所有的决策变量除一个外，对其余决策变量都固定在它们的标称值上.若有某个决策变量的变化对

后果有较大的影响，则把它包括到模型中来．若另一个决策变量的变化对后果只有微小的影响，甚至没有影响，则在灵敏度分析后设它为常数值，并从决策变量中去掉．对状态变量也可以作灵敏度分析，若有某个状态变量的变化对后果有较大的影响，此状态变量就是关键状态变量，是一个不确定变量，并在下阶段中，即随机阶段中，设定它的概率分布，否则，只采用它的标称值．

2．在随机阶段中，也要建立模型和对模型进行分析．建模中应包含设定概率分布和风险偏爱两个方面．分析是随机优势分析和灵敏度分析，还可以进行风险灵敏度分析．进行这种分析时，设效用函数是指数函数，用风险厌恶常数（它的倒数称为风险容限系数）去表示决策者对风险厌恶的程度．

3．在信息阶段中，先分析，后模型化．前者应计算完全信息期望值或计算采样信息期望值，更精确的是进行序贯分析．分析收集信息的必要性和可能性．收集新信息后，应设定关键变量的后验概率并研究修改模型结构的必要性．若模型结构有改变，则重复进行确定阶段和随机阶段的分析．

价值模型（value model）　一类特殊决策模型．分为开环价值模型和闭环价值模型两种，用它去计算各种后果的价值．“开环”是指这种价值完全是客观的，不包含决策者的主观意图．“闭环”是指这种价值包含决策者的意图．

多目标决策（multiple criteria decision making）　亦称多指标决策．决策论的一个新的分支．是研究含有多于一个目标（指标）的决策问题的学科领域．对于目标能定量描述，并且问题用极小化或极大化形式表示的情况，即指多目标规划（参见“多目标规划”）；对于目标只能定性描述，并且供选方案为有限个的情况，常称多属性决策．

多目标决策问题（multiple criteria decision making problem）　亦称多指标决策问题．一类决策问题．指包含有多于一个目标（指标，准则）的决策问题．它包括决策单元、目标、属性、决策情况和决策规则．决策单元包含决策者、分析者和其他，他们结合起来，做出决策．最小的决策单元就是决策者本人．目标是关于被研究问题的某个决策者所希望达到的状态的陈述．在一个多目标决策问题中，有若干种陈述表达决策者所希望达到的状态．目标可表示为一个递阶结构，它的最高层是总体目标，第二层是次目标，第三层是分目标等．在递阶结构中，下层的目标比上层的目标更加明确、更加具体和更便于计算．对于最下层每个目标的设定要满足可测性，并反映特定目标达到目的的程度．每个目标都必须满足可理解性和可测性．属性是指有的目标并没有一个或若干个明显的指标去直接测量它所达到的程度，但仍然可能存在一个或若干个属性，既便于测量又能间接地反映目标达到的程度．这种属性称为代用属性．决策情况是指这个决策问题的结构和决策环境．在不确定性的环境状态中包含已知概率分布和未知概率分布．决策规则是用于排列方案的优劣次序的规则，包括最优规则和满意规则两种．

多目标决策问题有目标间的不可公度性和目标间的矛盾性．求解多目标决策问题，通常通过以下四个步骤来完成：

1．提出目标．决策者分析各种情况，提出需要达到的目标．

2．构成问题．即明确问题，标明目的和辨识属性．

3．构造模型．这个模型包含若干决策变量和它们之间的关系，并估计参数．

4．分析和评价．首先建立供选方案集，然后估计各目标属性的值．

为了对方案进行评价和分析，预先规定决策规则，把每个方案和其他方案作比较，然后排列所有方案的优劣次序，做出决定，进行实施．若过程到此终止，则它是一个“开环”过程．当得到的结果还不能满意时，可以使用被选中的方案所得到的信息，再去构造多目标决策问题，从而形成一“闭环”过程．在第 2 步、第 3 步和第 4 步，还需进行价值判断．“判断”是在制定决策的过程中的价值元素．价值元素的集合，构成价值系统．

多目标决策规则（multicriteria decision rule）　一种特殊的决策规则．指对多目标决策问题进行决策的规则．通常是：决策者按照多个目标的属性 f_1，f_2，…，f_m 的值（$m\geqslant 2$），在供选方案集 X 中选择一个最偏爱的方案．对于极小化问题，在确定情况下，常用的决策规则是：选择某方案 $\hat{x}$，它使各属性 f_1，f_2，…，f_m 的值依次不大于各给定的对应值 $\overset{0}{f}_1$，$\overset{0}{f}_2$，…，$\overset{0}{f}_m$．即求问题的非劣解．设 $\tilde{X}$ 是非劣解集，则从中选择决策者最偏爱的那个方案．选择决策者偏爱的满意方案依赖于决策者的偏爱结构．设决策者的偏爱结构能用一效用函数 $u(f_1,f_2,\cdots,f_m)$ 表示，则选择一决策者偏爱的非劣解，即求解数值极大化问题

$$\max_{x\in X} u(f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x})).$$

价值函数（value function）　反映集合中元素间序关系的函数．设 $\succcurlyeq$ 是在集合 X 上的一弱序，且 X 的无差异类的集 $\mathscr{X}$ 是可测的，则存在一实值函数 $v: X\to \mathrm{R}$，使对 X 中的任何 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 有：$\boldsymbol{x}\succcurlyeq\boldsymbol{y}$ 当且仅当 $v(\boldsymbol{x})\geqslant v(\boldsymbol{y})$，或者 $\boldsymbol{x}\succ\boldsymbol{y}$ 当且仅当 $v(\boldsymbol{x})>v(\boldsymbol{y})$，以及 $\boldsymbol{x}\sim\boldsymbol{y}$ 当且仅当 $v(\boldsymbol{x})=v(\boldsymbol{y})$．在构造价值函数时，对 $\mathscr{X}$ 中的每个元素设定一相应的 v 值，它表示

该元素排列的次序. i 个元素排序的目录记为 $L(i)$：

1. 令 $v(\boldsymbol{x}_1)=0, L(1)=\{1\}$.

2. 设 $\boldsymbol{x}_2 \succ \boldsymbol{x}_1$，可令 $v(\boldsymbol{x}_2)=1$，并令

$$L(2)=\{2,1\}.$$

3. 由 $\boldsymbol{x}_3 \succ \boldsymbol{x}_2 \succ \boldsymbol{x}_1$，可令 $v(\boldsymbol{x}_3)=2$，并令

$$L(3)=\{3,2,1\}.$$

4. 观察到 $\boldsymbol{x}_3 \succ \boldsymbol{x}_4 \succ \boldsymbol{x}_2$，令

$$v(\boldsymbol{x}_4)=\frac{1}{2}[v(\boldsymbol{x}_3)+v(\boldsymbol{x}_2)]=\frac{3}{2},$$

则 $L(4)=\{3,4,2,1\}$.

5. 观察到 $\boldsymbol{x}_5 \succ \boldsymbol{x}_3 \succ \boldsymbol{x}_4$，可令 $v(\boldsymbol{x}_5)=3$，则

$$L(5)=\{5,3,4,2,1\}.$$

6. 观察到 $\boldsymbol{x}_6 \succ \boldsymbol{x}_5 \succ \boldsymbol{x}_3$，可令 $v(\boldsymbol{x}_6)=4$，则

$$L(6)=\{6,5,3,4,2,1\}.$$

7. 观察到 $\boldsymbol{x}_3 \succ \boldsymbol{x}_7 \succ \boldsymbol{x}_4$，令

$$v(\boldsymbol{x}_7)=\frac{1}{2}[v(\boldsymbol{x}_3)+v(\boldsymbol{x}_4)]=\frac{7}{4},$$

则 $L(7)=\{6,5,3,7,4,2,1\}$.

8. 可以类似地继续排出 $\boldsymbol{x}_8, \boldsymbol{x}_9, \cdots$，并设定或计算相应的 v 值.

设 v 为在 X 上由优先序 $\succcurlyeq$ 所产生的一价值函数. 令 V 为定义在 $v(X)$ 上的一严格递增的实值变换（保序变换），即 $V(v_1)>V(v_2)$ 当且仅当 $v_1>v_2$；$V(v_1)=V(v_2)$ 当且仅当 $v_1=v_2$. 这样，对于任何 $\boldsymbol{x}, \boldsymbol{y} \in X$，有 $\boldsymbol{x} \succcurlyeq \boldsymbol{y}$ 当且仅当 $V^*(\boldsymbol{x}) \geqslant V^*(\boldsymbol{y})$，而 $V^*(\boldsymbol{x})=V(v(\boldsymbol{x}))$. 令 $X \subset \mathrm{R}^n$，$\succcurlyeq$ 是在 X 上的一弱序，再设：

1）对于任何 $\boldsymbol{x}, \boldsymbol{y} \in X, \boldsymbol{x} \succcurlyeq \boldsymbol{y}$，均有 $\boldsymbol{x} \succ \boldsymbol{y}$；

2）对于任何 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in X$，若 $\boldsymbol{x} \succ \boldsymbol{y} \succ \boldsymbol{z}$，必有一个 $\lambda \in (0,1)$，使 $\boldsymbol{y} \sim \lambda \boldsymbol{x}+(1-\lambda) \boldsymbol{z}$；

则存在一定义在 X 上的实值函数 v，适合：$\boldsymbol{x} \succcurlyeq \boldsymbol{y}$ 当且仅当 $v(\boldsymbol{x}) \geqslant v(\boldsymbol{y})$；$\boldsymbol{x} \succ \boldsymbol{y}$ 当且仅当 $v(\boldsymbol{x})>v(\boldsymbol{y})$. 设有一定义在集合 Y 上的价值函数 v，$v(\boldsymbol{y})$ 称为是可测的当且仅当 v 反映了在 Y 中各元素的顺序和在 Y 中的元素之差的顺序.

加性价值函数（additive value function） 一类特定的价值函数. 把多属性集合区分为较小的子集，这些子集可以当做互相独立的，且对每个属性构造一价值函数，每次构造一个，然后按加性的形式把它们组合起来. 称这样组合起来的价值函数为加性价值函数，亦称决策者的偏爱结构是加性的. 令 $Y_1, Y_2, \cdots, Y_m$ 分别表示 m 个属性的可解值的集合，且令 Y 为属性集 $Y_1, Y_2, \cdots, Y_m$ 的笛卡儿乘积集，即 $Y=Y_1 \times Y_2 \times \cdots \times Y_m$. 一特殊的方案是 $\boldsymbol{y}=(\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_m)$，而 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_m$ 各为 $Y_1, Y_2, \cdots, Y_m$ 中相应的值. 按照这种记法，如果一偏爱结构能用某个价值函数 v 表示，则它是加性的当且仅当

$$v(\boldsymbol{y})=v_1(\boldsymbol{y}_1)+v_2(\boldsymbol{y}_2)+\cdots+v_m(\boldsymbol{y}_m)$$

或

$$v(\boldsymbol{y})=k_1 v_1(\boldsymbol{y}_1)+k_2 v_2(\boldsymbol{y}_2)+\cdots+k_m v_m(\boldsymbol{y}_m)$$

$[k_i(i=1,2,\cdots,m)]$ 为标度常数，

$$\sum_{i=1}^{m} k_i=1.$$

加性定理（additive theorem） 关于加性价值函数性质的主要结论. 有两个属性的加性定理和 m（>2）个属性的加性定理. 在两个属性的决策问题中，设 Y_1 和 Y_2 是属性集，当对于任何 $x_1, y_1, a_1 \in Y_1$ 和 $x_2, y_2, a_2 \in Y_2$，有 $(x_1, a_2) \succcurlyeq (a_1, y_2)$ 和 $(a_1, x_2) \succcurlyeq (y_1, a_2)$，必有 $(x_1, x_2) \succcurlyeq (y_1, y_2)$，则称优先序 $\succcurlyeq$ 适合消去条件. 消去条件是加性偏爱结构的必要条件. 由 $\succcurlyeq$ 所描绘的偏爱结构是指

$$(y_1', y_2') \succcurlyeq (y_1'', y_2'') \Leftrightarrow v(y_1', y_2') \geqslant v(y_1'', y_2''),$$

于是有：

1. 如消去条件被满足，则 $\succcurlyeq$ 是互相偏爱独立的.

2. 在 Y_1 和 Y_2 上如存在 v_1 和 v_2，使对于在 $Y=Y_1 \times Y_2$ 上的任何 $\boldsymbol{x}=(x_1, x_2)$ 和 $\boldsymbol{y}=(y_1, y_2)$，$\boldsymbol{x} \succcurlyeq \boldsymbol{y} \Leftrightarrow v_1(x_1)+v_2(x_2) \geqslant v_1(y_1)+v_2(y_2)$ 当且仅当消去条件被满足.

对于具有 m 个属性的决策问题，设有一定义在 Y 上的价值函数 $v(\boldsymbol{y})=v(y_1, y_2, \cdots, y_m)$，对于任何 $\boldsymbol{y}', \boldsymbol{y}'' \in Y$，有 $\boldsymbol{y}' \succcurlyeq \boldsymbol{y}''$ 当且仅当 $v(\boldsymbol{y}') \geqslant v(\boldsymbol{y}'')$，则存在分别定义在 $Y_1, Y_2, \cdots, Y_m$ 上的实值函数 $v_1, v_2, \cdots, v_m$，使 $\boldsymbol{y}' \succcurlyeq \boldsymbol{y}''$ 和 $v_1(y_1')+v_2(y_2')+\cdots+v_m(y_m') \geqslant v_1(y_1'')+v_2(y_2'')+\cdots+v_m(y_m'')$ 当且仅当在这集合属性上互相偏爱独立条件成立. 若存在定义在 $Y_1, Y_2, \cdots, Y_m$ 上的实值函数 $v_1, v_2, \cdots, v_m$ 和定义在 Y 上的 v，对于任何 $\boldsymbol{y}', \boldsymbol{y}'' \in Y$，使 $\boldsymbol{y}' \succcurlyeq \boldsymbol{y}''$ 当且仅当

$$v[v_1(y_1'), v_2(y_2'), \cdots, v_m(y_m')] \geqslant v[v_1(y_1''), v_2(y_2''), \cdots, v_m(y_m'')],$$

则 m 个属性的偏爱结构是可分解的.

汤姆森条件（Thomsen condition） 加性偏爱结构的必要条件. 在两个属性的决策问题中，设 Y_1 和 Y_2 是属性集. 若对于任何 $x_1, y_1, a_1 \in Y_1$ 和 $x_2, y_2, a_2 \in Y_2$，有 $(x_1, a_2) \sim (a_1, y_2)$ 和 $(a_1, x_2) \sim (y_1, a_2)$，必有 $(x_1, x_2) \sim (y_1, y_2)$，则称偏爱结构（优先序）$\succsim$ 适合汤姆森条件. 因消去条件含有汤姆森条件，故汤姆森条件是加性偏爱结构的必要条件.

拟加性形式（quasi-additivity form） 价值函数的一种形式. 指规范化的可测价值函数表示为系数不同的若干个条件可测价值函数积之和的形式. 设在属性集 Y 上存在规范的可测价值函数 v，而 $Y=Y_1 \times Y_2 \times \cdots \times Y_m$，在 Y 中存在两个极端的元素 $\boldsymbol{y}^0$ 和 $\boldsymbol{y}^*$，使 $v(\boldsymbol{y}^0)=0$ 和 $v(\boldsymbol{y}^*)=1$. 若属性集 $Y_i(i=1,2,\cdots,m)$ 弱差独立于其补集，则存在定义在 Y_i 上的条件可测价值函数 v_i，对于任何 $\boldsymbol{y} \in Y$ 有拟加性形式：

$$v(\mathbf{y})=\sum_{i=1}^{m}k_iv_i(y_i)+\sum_{i=1}^{m}\sum_{j>i}^{m}k_{ij}v_i(y_i)v_j(y_j)$$
$$+\sum_{i=1}^{m}\sum_{j>i}^{m}\sum_{k>j}^{m}[k_{ijk}v_i(y_i)v_j(y_j)v_k(y_k)+\cdots$$
$$+k_{12\cdots m}v_1(y_1)v_2(y_2)\cdots v_m(y_m)].$$

对于$i=1,2,\cdots,m$，有$v_i(y_i^*)=1$和$v_i(y_i^0)=0$. 若所有$k_{ij},k_{ijk},\cdots,k_{12\cdots m}$均为零，则简化为一加性形式；若$k_{ij}=kk_ik_j,k_{ijk}=k^2k_ik_jk_k,\cdots$和

$$\sum_{i=1}^{m}k_i\neq 1,$$

则拟加性形式简化为

$$v(\mathbf{y})=\sum_{i=1}^{m}k_iv_i(y_i)+k\sum_{i=1}^{m}\sum_{j>i}^{m}k_ik_jv_i(y_i)v_j(y_j)$$
$$+k^2\sum_{i=1}^{m}\sum_{j>i}^{m}\sum_{k>j}^{m}[k_ik_jk_kv_i(y_i)v_j(y_j)v_k(y_k)+\cdots$$
$$+k^{m-1}k_1k_2\cdots k_mv_1(y_1)v_2(y_2)\cdots v_m(y_m)].$$

称此形式为乘法形式，它是拟加性形式的特殊情况. 若$\Omega=\{Y_1,Y_2,\cdots,Y_m\}$的每个子集弱差独立于其补集，则称属性集$\Omega$是互相弱差独立的. 设在$Y$上存在一规范的可测价值函数$v$，若属性集$\Omega$是互相弱差独立的，则$v$可表示为乘法形式.

加性效用函数(additivity utility function)　一类特殊的效用函数. 指函数

$$u(\mathbf{y})=\sum_{i=1}^{m}k_iu_i(y_i)+\sum_{i=1}^{m}\sum_{j>i}^{m}k_{ij}u_i(y_i)u_j(y_j)+\cdots$$
$$+\sum_{i=1}^{m}\sum_{j>i}^{m}\sum_{k>j}^{m}[k_{ijk}u_i(y_i)u_j(y_j)u_k(y_k)+\cdots$$
$$+k_{12\cdots m}u_1(y_1)u_2(y_2)\cdots u_m(y_m)],$$

式中u表示在Y上的规范化的效用函数，且$u(\mathbf{y}^*)=1,u(\mathbf{y}^0)=0,k_i,k_{ij},k_{ijk},\cdots(i,j,k=1,2,\cdots,m)$为标度常数，$u_i(y_i^*)=1,u_i(y_i^0)=0(i=1,2,\cdots,m)$.

决策矩阵(decision matrix)　由决策属性值表示的矩阵. 设$X=\{x_1,x_2,\cdots,x_n\}$表示供选方案集，$Y_i=(y_{i1},y_{i2},\cdots,y_{im})$表示第$i$个方案$x_i$的各属性值集，$y_{ij}$表示第$i$个方案第$j$个属性的值. 若用目标函数表示属性，则

$$y_{ij}=f_j(x_i)\quad(i=1,2,\cdots,n;j=1,2,\cdots,m).$$

各方案的属性值可用矩阵$(y_{ij})_{n\times m}$表示，称此矩阵为决策矩阵. 决策矩阵的各属性值y_{ij}可用规范化方法都统一变换到(0,1)范围内，记为z_{ij}. 常用的规范化方法有：

1. 向量规范化. 令

$$z_{ij}=\frac{y_{ij}}{\sqrt{\sum_{i=1}^{n}y_{ij}}},$$

此变换把所有属性值均化为量纲一的量，且均处于(0,1)范围内. 但此变换是非线性的，变换后各属性的最小值不为0，最大值不为1.

2. 线性变换. 若目标为效益，则令

$$y_{ij}=\frac{y_{ij}}{y_j^{\max}},$$

式中$y_j^{\max}=\max_i y_{ij}$. 据此，$y_j^{\max}$是决策矩阵第j列中的最大元素，于是有$0\leqslant z_{ij}\leqslant 1$. 若目标为成本，则令

$$z_{ij}=1-\frac{y_{ij}}{y_j^{\max}},$$

于是$0\leqslant z_{ij}\leqslant 1$. 此变换是线性的，变换后的相对数量和变换前的相同.

3. 其他变换. 在决策矩阵中既有效益目标又有成本目标，变换后最好的效益目标和最好的成本目标有不同的值，不便于比较. 把成本目标的变换修改为

$$z_{ij}=\frac{1/y_{ij}}{\max_i(1/y_{ij})}=\frac{\min_i y_{ij}}{y_{ij}}=\frac{y_i^{\min}}{y_{ij}},$$

式中的$y_i^{\min}$是决策矩阵第i行的最小元素. 对于效益目标，则令

$$z_{ij}=\frac{y_{ij}-y_j^{\min}}{y_j^{\max}-y_j^{\min}};$$

对于成本，则令

$$z_{ij}=\frac{y_j^{\max}-y_{ij}}{y_j^{\max}-y_j^{\min}}.$$

此变换可把属性的最大值统一为0和1，但此变换不是比例的.

筛选法(screening procedure)　一类决策方法. 有优选法、连结法(满意法)和分离法三种. 优选法是把一个方案和另一个方案作比较，淘汰那些劣的方案，找到比它们更好的方案，如此下去，总能淘汰掉一批劣方案. 连结法(满意法)是对每个属性提供一个能接受的最低值，称为切除值. 只有另一个方案的每个属性值均不低于对应的切除值时，该方案才不被删除，即方案能被接受. 若用迭代法逐步提高切除值，便能达到所希望保留的方案数. 分离法与连结法不同，不要求方案的每个属性值都超过切除值，只要求方案至少有一个属性值超过切除值就被保留.

简单加性加权法(weighted method of simple additivity)　求多属性决策问题的方法. 当属性集满足一定的独立性条件，则其价值(或效用)函数u都可以表示为加性形式，即

$$u(y_1,y_2,\cdots,y_m)=\sum_{i=1}^{m}k_iu_i(y_i),$$

式中的$u_i(i=1,2,\cdots,m)$表示第i个属性的价值(或效用)，y_i是第i个属性值，k_i是第i个属性的权重. 对上述加性加权形式的价值(或效用)函数求极大化，即为简单加性加权法. 若将效用u_i作近似线性表示，用规范化后的属性值z_i去代替y_i，则

$$u_i(z_i)=\mu_i z_i \quad (i=1,2,\cdots,m),$$

式中的 μ_i 为常数. 令 $k_i\mu_i=w_i$,则有

$$u(z_1,z_2,\cdots,z_m)=\sum_{i=1}^{m}w_i z_i.$$

层次加性加权法(weighted method of stratification additivity) 一种分层决策的方法. 把一个有限方案的多目标决策问题分解为若干层次,逐层用简单加性加权法进行求解的方法.

逼近理想解排序法(technique for sequencing by approximate ideal solution or technique for order preference by similarity to ideal solution) 一种借助于多属性决策问题的理想解和负理想解,对多属性决策问题的供选方案进行排序的方法. 理想解是设想的最好解,它使各属性值都达到各供选方案中的最好值;负理想解是设想的最坏解,它的各属性值都达到各供选方案中的最坏值. 设研究 n 个供选方案和 m 个属性的多属性决策问题,当采用欧几里得范数作为距离的测度时,解 $\boldsymbol{x}^i$ 到理想解 $\boldsymbol{x}^*$ 的距离是

$$S_i^*=\sqrt{\sum_{j=1}^{m}(x_{ij}-x_j^*)^2} \quad (i=1,2,\cdots,n),$$

式中 x_{ij}是解 $\boldsymbol{x}^i$ 的第 j 个分量,即第 j 个属性的规范化的加权值,x_j^* 是理想解 $\boldsymbol{x}^*$ 的第 j 个分量. 类似地,解 $\boldsymbol{x}^i$ 到负理想解 $\boldsymbol{x}^-$ 的距离是

$$S_i^-=\sqrt{\sum_{j=1}^{m}(x_{ij}-x_j^-)^2} \quad (i=1,2,\cdots,n),$$

解 $\boldsymbol{x}^i$ 对理想解的相对接近度为

$$c_i^*=\frac{S_i^-}{S_i^*+S_i^-} \quad (0\leqslant c_i^*\leqslant 1,\ i=1,2,\cdots,n).$$

若 $\boldsymbol{x}^i$ 为理想解 $\boldsymbol{x}^*$,则 c_i^* 为 1;c_i^* 越接近于 1,方案越排在前面. 若 $\boldsymbol{x}^i$ 为负理想解 $\boldsymbol{x}^-$,则 $c_i^*=0$. 按 c_i^* 由大到小的顺序排列,排在前面的方案应优先采用.

线性分配法(linear assignment method) 一种对多属性决策问题进行方案排序的方法. 该法不先给出决策矩阵,只知各供选方案对每个属性的优先次序,该方法需构造一个权矩阵去估计每个方案排列的位置. 设有 n 个供选方案和 m 个属性,先根据给定的权 $(w_1,w_2,\cdots,w_m)^{\mathrm{T}}$ 和方案对每个属性的排序,求得一个 $n\times n$ 权矩阵 $W=(w_{ik})_{n\times n}$. 再定义一排列矩阵 P,它的元素 p_{ik}为 0 或 1. $p_{ik}=1$ 表示方案 x_i 被排在第 k 位. 若 $p_{ik}=1(i=1,2,\cdots,n)$,则

$$p_{ij}=0 \quad (j=1,2,\cdots,n;j\neq k).$$

若 $p_{ik}=1(k=1,2,\cdots,n)$,则

$$p_{jk}=0 \quad (j=1,2,\cdots,n;j\neq i).$$

由此,线性分配方案归为求解线性规划问题

$$\begin{cases}\max & \displaystyle\sum_{i=1}^{n}\sum_{k=1}^{n}w_{ik}p_{ik},\\ \text{s.t.} & \displaystyle\sum_{k=1}^{n}p_{ik}=1 \quad (i=1,2,\cdots,n),\\ & \displaystyle\sum_{i=1}^{n}p_{ik}=1 \quad (k=1,2,\cdots,n).\end{cases}$$

求解此线性规划问题,即求一组 $p_{ik}=1$ 的值,使对应的权的和达到最大.

相对位置估计法(estimate method of relative place) 一种序数决策方法. 其步骤如下:

1. 估计偏爱(优先)关系. 根据决策者对供选方案 $\boldsymbol{x}_i(i=1,2,\cdots,n)$的后果的了解和偏爱,估计方案集中元素的偏爱(优先)关系,偏爱关系可以用指向图表示,还可以用 $n\times n$ 阶矩阵 M 表示. 设 n 为方案数,关于矩阵 M 中的元素 $m_{ij}(i,j=1,2,\cdots,n)$,如 $\boldsymbol{x}_i\succ\boldsymbol{x}_j$ 或 $\boldsymbol{x}_i\sim\boldsymbol{x}_j$,取值 1;如 $\boldsymbol{x}_j\succ\boldsymbol{x}_i$(或者方案 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 不能比较)取值 0. 不论用何种方法,都要确定其偏爱关系 $\boldsymbol{x}_i\succ\boldsymbol{x}_j$ 或 $\boldsymbol{x}_j\succ\boldsymbol{x}_i$,或 $\boldsymbol{x}_i\sim\boldsymbol{x}_j$.

2. 构成排序指示值. 对每个方案 $\boldsymbol{x}_i$ 确定相对位置指数 $v(\boldsymbol{x}_i)=p(\boldsymbol{x}_i)-q(\boldsymbol{x}_i)$,$p(\boldsymbol{x}_i)=|\{\boldsymbol{x}_j|\boldsymbol{x}_j\in X,\boldsymbol{x}_j\neq\boldsymbol{x}_i,\boldsymbol{x}_i\succ\boldsymbol{x}_j,\text{非 }\boldsymbol{x}_j\succ\boldsymbol{x}_i\}|$ 是 X 中比 $\boldsymbol{x}_i$ 差的方案的数目,$q(\boldsymbol{x}_i)=|\{\boldsymbol{x}_j|\boldsymbol{x}_j\in X,\boldsymbol{x}_j\neq\boldsymbol{x}_i,\boldsymbol{x}_j\succ x_i,\text{非 }\boldsymbol{x}_i\succ\boldsymbol{x}_j\}|$ 是 X 中比 $\boldsymbol{x}_i$ 好的方案的数目.

3. 用排序指示值排序. 用 v 的值对 X 中的诸方案排序. 最大 v 值的方案排在第一位,次大的排在第二位等. 如两元素的 v 相等,则位置相同.

在多目标情况下,为了判断两个方案的优劣,需要对各目标加权,即用 $w_j(j=1,2,\cdots,m)$记第 j 个目标 f_j 的权,且令

$$\sum_{j=1}^{m}w_j=1.$$

若对于第 j 个目标 f_j 有 $\boldsymbol{x}_i\succ\boldsymbol{x}_k$,则构成一集$(\boldsymbol{x}_i\succ\boldsymbol{x}_k)$. 若 m 个目标中有若干个 $\boldsymbol{x}_i\succ\boldsymbol{x}_k$,则有

$$w(\boldsymbol{x}_i\succ\boldsymbol{x}_k)=\sum_{\boldsymbol{x}_i\succ\boldsymbol{x}_k}w_j.$$

同样,可以定义 $w(\boldsymbol{x}_k\succ\boldsymbol{x}_i)$和 $w(\boldsymbol{x}_i\sim\boldsymbol{x}_k)$. 设 $A_\sigma(\boldsymbol{x}_i\succ\boldsymbol{x}_k)$为指示值,表示方案 $\boldsymbol{x}_i\succ\boldsymbol{x}_k$ 的偏爱程度,σ 为一非负实数,则有

$$A_\sigma(\boldsymbol{x}_i\succ\boldsymbol{x}_k)=\frac{w(\boldsymbol{x}_i\succ\boldsymbol{x}_k)+\sigma w(\boldsymbol{x}_i\sim\boldsymbol{x}_k)}{w(\boldsymbol{x}_i\prec\boldsymbol{x}_k)+\sigma w(\boldsymbol{x}_i\sim\boldsymbol{x}_k)}.$$

使用这一指示值,可推出在 m 个目标下,两个方案之间的偏爱(优先)关系. 即若 $A_\sigma(\boldsymbol{x}_i\succ\boldsymbol{x}_k)>A'$,则 $\boldsymbol{x}_i\succ\boldsymbol{x}_k$;若 $A_\sigma(\boldsymbol{x}_i\succ\boldsymbol{x}_k)<A''$,则 $\boldsymbol{x}_i\prec\boldsymbol{x}_k$;若 $A''\leqslant A_\sigma(\boldsymbol{x}_i\succ\boldsymbol{x}_k)\leqslant A'$,则 $\boldsymbol{x}_i\sim\boldsymbol{x}_k$,式中的 $A'(\geqslant 1)$和 $A''(\leqslant 1)$为某固定的阀值. 若 $w(\boldsymbol{x}_i\succ\boldsymbol{x}_k)+\sigma w(\boldsymbol{x}_i\sim\boldsymbol{x}_k)$比 $w(\boldsymbol{x}_i\prec\boldsymbol{x}_k)+\sigma w(\boldsymbol{x}_i\sim\boldsymbol{x}_k)$大,则 $\boldsymbol{x}_i\succ\boldsymbol{x}_k$;反之,则 $\boldsymbol{x}_k\succ\boldsymbol{x}_i$. 若 A_σ 的值在 1 的附近,则 $\boldsymbol{x}_i\succ\boldsymbol{x}_k$ 的权因子和 $\boldsymbol{x}_k\succ\boldsymbol{x}_i$ 的权因子几乎相等,故 $\boldsymbol{x}_i\sim\boldsymbol{x}_k$. 若 $\sigma=0$,则 A_σ 可以忽略不

计，若 σ 很大，则 A_σ 受到这些几乎无差异的目标的影响，从而使两个方案 $x_i \sim x_k$. 因为

$$A_\sigma(x_i \succ x_k) = \frac{1}{A_\sigma(x_k \prec x_i)},$$

故阈值 A''可取为 $1/A'$.

选择法(choice method or elimination et choice translating reality method)　一种基于认可与否的多属性决策方法. 是对于供选方案间的偏爱关系无法确认，而只是决策者觉得有理由相信大致上认可有某种偏爱关系，并愿意承担由这种"认可"带来的风险时，解决该类多属性决策问题的有效方法. 算法如下：

1. 对供选方案集 $X=\{x_1,x_2,\cdots,x_n\}$ 中的每个方案 $x_i(i=1,2,\cdots,n)$ 依次确定相应的属性 $y_1(x_i)$, $y_2(x_i),\cdots,y_m(x_i)$ 的值 $y_{i1},y_{i2},\cdots,y_{im}$，并用矩阵 $A=(y_{ij})_{n\times m}$ 表示.

2. 由和谐性检验和不和谐性检验构造级别高于关系，并给出这种级别高于关系的指向图. 级别高于关系 S 和级别无差异关系 $\bar{S}$ 都是二元关系，它们的基本性质是：对于 X 中的任何方案 x',x^0 和 x''，有 $x'Sx^0$和$y(x^0)\geqslant y(x'')\Rightarrow x'Sx''$，以及 $y(x')\geqslant y(x^0)$ 和 $x^0Sx''\Rightarrow x'Sx''$. 这个性质称为弱传递性. 由弱传递性，可得$y(x')\geqslant y(x'')\Rightarrow x'Sx''$. $\bar{S}$ 是自反的、对称的和传递的. 构造 S 的方法是和谐性与不和谐性的两次检验. 第一次称为和谐性检验，第二次称为不和谐性检验.

和谐性检验有以下步骤：

1) 决策者要求指出属性 $y_1,y_2,\cdots,y_m$ 的相对重要性，并用一组权 $w_1,w_2,\cdots,w_m$ 表示其相对重要性.

2) 对于 X 中的每对方案 x_j 和 x_k 构造以下指标集：$I^+(x_j,x_k)=\{i\mid 1\leqslant i\leqslant m,y_i(x_j)>y_i(x_k)\}$, $I^=(x_j,x_k)=\{i\mid 1\leqslant i\leqslant m,y_i(x_j)=y_i(x_k)\}$, $I^-(x_j,x_k)=\{i\mid 1\leqslant i\leqslant m,y_i(x_j)<y_i(x_k)\}$.

3) 构造和谐性指数 I_{jk}和 $\hat{I}_{jk}$，即

$$I_{jk}=\Big(\sum_{i\in I^+(x_j,x_k)} w_i+\sum_{i\in I^=(x_j,x_k)} w_i\Big)\Big/\Big(\sum_{i=1}^{m} w_i\Big)$$

和

$$\hat{I}_{jk}=\Big(\sum_{i\in I^+(x_j,x_k)} w_i\Big)\Big/\Big(\sum_{i\in I^-(x_j,x_k)} w_i\Big).$$

4) 选择和谐性最低值参数 α，若 $I_{jk}\geqslant\alpha,\hat{I}_{jk}\geqslant 1$，则认为通过了和谐性检验.

不和谐性检验有以下步骤：

1) 设定权 $w_1,w_2,\cdots,w_m$，并构造 $I^-(x_j,x_k)$.

2) 对每个 $i=1,2,\cdots,n$，构造不和谐集 $D_i=\{(y_i(x_j),y_i(x_k))\mid y_i(x_j),y_i(x_k)\in Y_i,y_i(x_k)-y_i(x_j)\geqslant d_i;j,k=1,2,\cdots,n;j\neq k\}(i=1,2,\cdots,m)$. 因此，对于两个方案 x_j 和 x_k，若 $y_i(x_k)-y_i(x_j)\geqslant d_i$，则不论其他属性如何，不承认 x_jSx_k. 这样，d_i 可以当做极限，在它之内补偿是允许的，在它之外不能补偿.

3) 对于一对给定的方案 x_j 和 x_k，若对于所有 $i\in I^-(x_j,x_k)$有$(y_i(x_j),y_i(x_k))\notin D_i$，则不和谐性检验被满足. 因此，如仅使用和谐性检验，则 x_jSx_k 当且仅当 $I_{jk}\geqslant\alpha$ 和 $\hat{I}_{jk}\geqslant 1$ 成立. 若不仅使用和谐性检验，还使用不和谐性检验，则 x_jSx_k 当且仅当 $I_{jk}\geqslant\alpha$, $\hat{I}_{jk}\geqslant 1$ 和 $((y_i(x_j),y_i(x_k))\notin D_i$，对于所有 $i\in I^-(x_j,x_k)$两式都成立.

3. 由 X 得到最小优势子集 X_1，如有核存在，则选择此核为 X_1. 若去掉某些或所有这种级别较低的方案，再去掉回路中至少一个方案，则余下的方案称为 X 相对于 S 的最小优势子集 X_1. 最小优势子集的性质是：对于每个 $x'\in\bar{X}_1(\bar{X}_1$ 是 X_1 的补集)有 x^*Sx'，对于某些 $x^*\in X_1$，最小优势子集 X_1 不是惟一的. 若最小优势子集 X_1 中没有两个方案可以由 S 进行比较，即在 X_1 中没有一对节点直接由指向连结，则这种优势子集称为 X 的核. 如 X 有核存在，必须把它用到下一步. 若无核，只能从最小优势子集中任选一个.

4. 如 X_1 只有单个方案，或者方案数目少到可以使用价值判断去选择最终方案，则过程终止，否则，修改 α 或 d_i 的值，重复步骤 2 到步骤 4.

修正选择法(modified choice method or modified elimination et choice translating reality method)　一类基于认可与否的多属性决策方法. 它弥补了选择法中使用阈值 α 和 d_i 的缺点(参见"选择法"). 对于 $I_{jk}\geqslant\alpha,\hat{I}_{jk}\geqslant 1$ 的和谐性的最低值定义了三个递减的阀值：α^*,α^0 和 α^-，并依次称为高、中和低的阀值，要求 $1\geqslant\alpha^*\geqslant\alpha^0\geqslant\alpha^-\geqslant 0.5$. 定义三个不和谐集：

低不和谐集

$$D_i^{\mathrm{l}}=\{(y_i(x_j),y_i(x_k)\mid y_i(x_k)-y_i(x_j)\leqslant\alpha_i^0,j,k=1,2,\cdots,m,j\neq k\},$$

中不和谐集

$$D_i^{\mathrm{m}}=\{(y_i(x_j),y_i(x_k))\mid d_i^0<y_i(x_k)-y_i(x_j)\leqslant d_i^*,j,k=1,2,\cdots,m,j\neq k\},$$

高不和谐集

$$D_i^{\mathrm{h}}=\{(y_i(x_j),y_i(x_k))\mid d_i^*<y_i(x_k)-y_i(x_j),j,k=1,2,\cdots,m,j\neq k\}.$$

方案的强级别高于关系和弱级别高于关系是：

1. 方案 x_j 强级别高于方案 x_k：若和谐性高，而不和谐性中；或者和谐性中，而不和谐性低. 用 S_p 记"强级别高于"，上述定义可记为：$x_jS_px_k\Leftrightarrow\hat{I}_{jk}\geqslant 1$，以及如果 $I_{jk}\geqslant\alpha^*$ 和 $(y_i(x_j),y_i(x_k))\in D_i^{\mathrm{m}}$, $i\in I^-(x_j,x_k)$；或者如果 $I_{jk}\geqslant\alpha^0$ 和$(y_i(x_j),y_i(x_k))\in$

$D_i^l, i \in I^-(\boldsymbol{x}_j, \boldsymbol{x}_k)$.

2. 方案 $\boldsymbol{x}_j$ 弱级别高于方案 $\boldsymbol{x}_k$：若和谐性及不和谐性都是低的；或者和谐性及不和谐性都是中的. 用 S_j 记"弱级别高于"，则上述定义可记为 $\boldsymbol{x}_j S_j \boldsymbol{x}_k \Leftrightarrow \hat{I}_{jk} \geqslant 1$，以及若 $I_{jk} \geqslant \alpha^-$ 和 $(y_i(\boldsymbol{x}_j), y_i(\boldsymbol{x}_k)) \in D_i^l, i \in I^-(\boldsymbol{x}_j, \boldsymbol{x}_k)$；或者若 $I_{jk} \geqslant \alpha^0$ 和 $(y_i(\boldsymbol{x}_j), y_i(\boldsymbol{x}_k)) \in D_i^m$，$i \in I^-(\boldsymbol{x}_j, \boldsymbol{x}_k)$.

由二元关系 S_p 和 S_j，可对方案集 X 做出两个指向图，图中的节点表示方案，而节点间的枝所指方向表示两个方案间强的（或弱的）关系. 用 G_F 和 G_j 分别记强图和弱图，用 Y 记图中节点数，用 U_F 和 U_j 分别记强图和弱图中的数，则强图为 $G_F(Y, U_F)$，弱图为 $G_j(Y, U_j)$. 此法排序步骤有前向排序、反向排序两种，并由此得到平均序.

多维偏爱分析的线性规划法(linear programming techniques for multidimensional analysis of preference)　借助于理想解评价方案优劣的一种多目标(属性)决策方法. 用 m 维目标空间中的 n 个点表示一个有 m 个属性的多目标决策问题的 n 个供选方案，并设想决策者所最偏爱的方案能用目标空间中的一理想点表示（参见"理想点法"）. 为了比较两个方案靠近理想点的程度，定义空间中任一点 $(z_{i1}, z_{i2}, \cdots, z_{im})$ 到理想点 $\boldsymbol{z}^* = (z_1^*, z_2^*, \cdots, z_m^*)$ 的加权欧几里得距离为

$$d_i = \left[\sum_{j=1}^{m} w_j (z_{ij} - z_j^*)^2\right]^{\frac{1}{2}} \quad (i = 1, 2, \cdots, n),$$

式中的 w_j 是第 j 个属性的权，平方距离为

$$S_i = d_i^2 = \sum_{j=1}^{m} w_i (z_{ij} - z_i^*)^2 \quad (i = 1, 2, \cdots, n),$$

式中的权向量 $\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^{\mathrm{T}}$ 和理想点的位置都是待确定的. 分析者选择了权向量和理想点的值 $(\boldsymbol{w}, \boldsymbol{z}^*)$ 后，计算每对方案 (k, l) 和理想点间的加权欧几里得距离：

$$S_k = \sum_{j=1}^{m} w_j (z_{kj} - w_j^*)^2$$

和

$$S_l = \sum_{j=1}^{m} w_j (z_{lj} - z_j^*)^2,$$

式中的 z_{kj} 是第 k 个方案 $\boldsymbol{x}_k$ 的第 j 个属性的值，z_{lj} 是第 l 个方案 $\boldsymbol{x}_l$ 的第 j 个属性的值. 对于 $\Omega = \{(k, l)\}$ 中的一对方案 (k, l)，若相应的 S_k 和 S_l 为 $S_l \geqslant S_k$，则方案 k 较方案 l 更靠近或同样靠近理想点，因此加权距离模型将和决策者的偏爱一致. 反之，若 $S_l < S_k$，则加权距离模型和决策者的偏爱不一致. 权向量和理想点 $(\boldsymbol{w}, \boldsymbol{z}^*)$ 的值的选择应使加权距离模型尽量和决策者的偏爱一致，尽量符合有序对 (k, l) 和集 Ω. 设用

$$(S_l - S_k)^- = \begin{cases} 0 & (S_l \geqslant S_k), \\ S_k - S_l & (S_l < S_k) \end{cases}$$
$$= \max\{0, \ (S_k - S_l)\}$$

度量加权距离模型和有序对 (k, l) 不一致的程度. 若 $S_l \geqslant S_k$，则加权距离模型和有序对 (k, l) 一致，作为不一致程度的量 $(S_l - S_k)^-$ 应为零. 如 $S_l < S_k$，则加权距离模型和有序对 (k, l) 不一致，S_k 和 S_l 的差愈大，不一致的程度愈高，因此 $(S_l - S_k)^-$ 应为 $S_k - S_l$. 对 Ω 中所有的 (k, l) 求和，得到一个量 B，表示加权距离模型和决策者的偏爱的总的不一致程度，称 B 为不一致度，即

$$B = \sum_{(k,l) \in \Omega} (S_l - S_k)^-.$$

设

$$G = \sum_{(k,l) \in \Omega} (S_l - S_k)^+,$$

其中

$$(S_l - S_k)^+ = \begin{cases} S_l - S_k & (S_l \geqslant S_k), \\ 0 & (S_l < S_k). \end{cases}$$

G 是一测度，反映了加权距离模型和决策者的偏爱的一致程度. 为了求得 B 的极小，添加约束条件 $G > B$ 或 $G - B = h$，h 是分析者所选择的某个正数. 由 $(S_l - S_k)^+$ 和 $(S_l - S_k)^-$，得到

$$(S_l - S_k)^+ - (S_l - S_k)^- = S_l - S_k,$$

$$G - B = \sum_{(k,l) \in \Omega} (S_l - S_k)^+ - \sum_{(k,l) \in \Omega} (S_l - S_k)^- = \sum_{(k,l) \in \Omega} (S_l - S_k) = h.$$

$(\boldsymbol{w}, \boldsymbol{z}^*)$ 可以通过求解约束最优化问题

$$\begin{cases} \min\left\{B = \sum_{(k,l) \in \Omega} \max\{0, (S_k - S_l)\}\right\}, \\ \text{s.t.} \ \sum_{(k,l) \in \Omega} (S_l - S_k) = h \end{cases}$$

得到. 这个约束最优化问题可变换为如下的线性规划问题：

$$\begin{cases} \min \sum_{(k,l) \in \Omega} \lambda_{kl}, \\ \text{s.t} \ (S_l - S_k) + \lambda_{kl} \geqslant 0 \ \ ((k, l) \in \Omega), \\ \quad \sum_{(k,l) \in \Omega} (S_l - S_k) = h, \\ \quad \lambda_{kl} \geqslant 0 \quad ((k, l) \in \Omega). \end{cases}$$

由加权欧几里得距离得到

$$S_l - S_k = \sum_{j=1}^{m} w_j (z_{lj} - z_j^*)^2 - \sum_{j=1}^{m} w_j (z_{kj} - z_j^*)^2$$
$$= \sum_{j=1}^{m} w_j (z_{lj}^2 - z_{kj}^2) - 2\sum_{j=1}^{m} w_j z_j^* (z_{lj} - z_{kj}).$$

用 v_j 代替 $w_j z_j^*$，上述线性规划模型成为

$$\begin{cases}\min \sum_{(k,l)\in\Omega} \lambda_{kj}, \\ \text{s.t.} \sum_{j=1}^{m} w_j(z_{lj}^2 - z_{kj}^2) \\ \quad - 2\sum_{j=1}^{m} v_j(z_{lj} - z_{kj}) + \lambda_{kj} \geqslant 0 \\ \quad (\forall (k,l) \in \Omega), \\ \quad \sum_{j=1}^{m} w_j \sum_{(k,l)\in\Omega} (z_{lj}^2 - z_{kj}^2) \\ \quad - 2\sum_{j=1}^{m} v_j \sum_{(k,l)\in\Omega} (z_{lj} - z_{kj}) = h, \\ \quad w_j \geqslant 0 \quad (j = 1,2,\cdots,m), \\ \quad \lambda_{kj} \geqslant 0 \quad ((k,l) \in \Omega).\end{cases}$$

群体决策(group decision making) 亦称群决策、集体决策、团体决策.决策论的一个重要分支.是研究如何集结决策群体中各决策个体的偏爱选择,以形成群体的偏爱,从而对供选方案进行群体偏爱排序或选优的学科.关于群体决策的研究,可以追溯到 1781 年鲍达(Borda,J.G.)对于选举制度的探索.群体决策研究的进展,与西方国家中福利经济学的发展密切相关.西方福利经济学派追求使社会中的资源和商品分配能产生最大的社会福利,并认为据此可以判断一种社会状况是优于、无差异于或是劣于另一种社会状况.1938 年,伯格森(Bergson,A.)在社会选择理论中引进社会福利函数,给群体决策提供了重要的表述形式.1951 年,阿罗(Arrow,K.J.)的《社会选择与个人价值》一书的出版,为群体决策的序数偏爱分析理论奠定了基础.20 世纪 70 年代以后,一些学者又发展了群体决策的基数偏爱分析理论和方法.

群体决策问题(group decision making problem) 一种决策问题.指需由多(于 1)个人共同进行决策的问题.参与决策的人称为决策个体或决策者,由决策个体组成的决策组称为决策群体,可供决策选择的对象称为供选方案.例如,需由董事会做出决定的有关企业的重大决策问题、由评判组进行打分评比的文体竞技活动,以及由全体选民参加投票的选举问题,都是群体决策问题.一个企业的董事会、文体竞赛的评判组、参加投票的全体选民,是相应群体决策问题的决策群体.企业决策中可供选择的方案、竞技比赛中的参赛者、选举活动中的候选人,则是相应问题的供选方案.

决策个体(decision maker) 亦称决策者.见“群体决策问题”.

决策群体(group of decision making) 有时简称群体.见“群体决策问题”.

供选方案(alternatives) 亦称备选方案(参见“群体决策问题”).由群体决策问题的所有供选方案组成的集合,称为该群体决策问题的供选方案集.

偏爱关系(preference relation) 集合上的一种二元关系.设 R 是集合 S 上的二元关系.对于任意 $a,b\in S$,若有 $aRb\Rightarrow bRa$ 或 $\neg bRa$(非 bRa),则称 R 是 S 上的偏爱关系或偏爱,aRb 读做 a 偏爱于 b;若有 $aRb\Rightarrow\neg bRa$,则称 R 是 S 上的严格偏爱(关系),记为 aPb,读做 a 严格偏爱于 b;若有 $aRb\Rightarrow bRa$,则称 R 是 S 上的淡漠(关系),记为 aIb,读做 a 淡漠于 b.有时,“严格偏爱于”和“淡漠于”也分别称为“优于”和“无差异于”.

群体偏爱映射(group preference mapping) 亦称群体一致函数.由各决策个体的偏爱到决策群体的偏爱的映射(参见“偏爱关系”).设 X 是供选方案集,$R_r(r=1,2,\cdots,l,l\geqslant 2)$ 是第 r 个决策个体 DM_r 在 X 上的偏爱关系,R 是决策群体 $G=\{DM_1,DM_2,\cdots,DM_l\}$ 在 X 上的偏爱关系,则映射 $F:\{R_1,R_2,\cdots,R_l\}\to R$ 称为 G 在 X 上的群体偏爱映射,记为 $R=F(R_1,R_2,\cdots,R_l)$.在福利经济学中,称 $F(R_1,R_2,\cdots,R_l)$ 为社会福利函数.

社会福利函数(social welfare function) 见“群体偏爱映射”.

阿罗不可能性定理(Arrow can not theorem) 群体决策中关于序数偏爱分析理论的基本定理.1951 年,阿罗(Arrow,K.J.)在《社会选择与个人价值》的著作中,提出合理的群体偏爱映射(他称之为社会福利函数)应满足:所谓完全性和传递性假设,以及一致性、独立性、非强加性和非独裁性等四个公理.但是,阿罗证明:在决策个体数不少于 2 和供选方案不少于 3 的条件下,不存在一群体偏爱映射,它同时能够满足上述两个假设和四个公理.这即阿罗不可能性定理.20 世纪 60 年代,关于阿罗的这一论断曾引起一场学术争论.以伯格森(Bergson,A.)为首的一些福利经济学家,对此持不同的看法.经过广泛的论战,现在一般认为:在序数偏爱的意义下,阿罗不可能性定理是正确的;但在基数偏爱的意义下,则可以存在群体偏爱映射满足相当于阿罗的假设和公理.

偏差度(preference-deviation measure) 关于偏爱的一种度量.是群体决策中,决策个体或决策群体关于两个供选方案偏爱之差的度量.设 X 是供选方案集,$\boldsymbol{x},\boldsymbol{y}\in X$.设 $R_r(r=1,2,\cdots,l)$ 是决策个体 DM_r 在 X 上的偏爱关系.若由 DM_r 确定的实数 $\theta_r(\boldsymbol{x},\boldsymbol{y})\in[-d,d](d>0)$ 满足:

1. $\theta_r(\boldsymbol{x},\boldsymbol{y})\geqslant 0\Leftrightarrow \boldsymbol{x}R_r\boldsymbol{y}$;
2. $\theta_r(\boldsymbol{x},\boldsymbol{y})+\theta_r(\boldsymbol{y},\boldsymbol{x})=0$;
3. 对任意的 $z\in X$,有

$$\theta_r(\boldsymbol{x},\boldsymbol{z}) + \theta_r(\boldsymbol{z},\boldsymbol{y}) = \theta_r(\boldsymbol{x},\boldsymbol{y});$$

则称 $\theta_r(\boldsymbol{x},\boldsymbol{y})$是决策个体 DM_r 关于供选方案对$(\boldsymbol{x},\boldsymbol{y})$的个体偏差度.

设 R 是决策群体 G 在 X 上的偏爱关系.若实数$\theta(\boldsymbol{x},\boldsymbol{y})$满足 $\theta(\boldsymbol{x},\boldsymbol{y})\geqslant 0 \Leftrightarrow \boldsymbol{x}R\boldsymbol{y}$,则称 $\theta(\boldsymbol{x},\boldsymbol{y})$是决策群体 G 关于供选方案对$(\boldsymbol{x},\boldsymbol{y})$的群体偏差度.

偏比度(preference-rate measure)　关于偏爱的一种度量.是群体决策中,决策个体或决策群体关于两个供选方案偏爱之比的度量.设 X 是供选方案集,$\boldsymbol{x},\boldsymbol{y}\in X$.

设 $R_r(r=1,2,\cdots,l)$是决策个体 DM_r 在 X 上的偏爱关系.若由 DM_r 确定的实数 $\varphi_r(\boldsymbol{x},\boldsymbol{y})>0$ 满足:

1. $\varphi_r(\boldsymbol{x},\boldsymbol{y})\geqslant 1 \Leftrightarrow \boldsymbol{x}R_r\boldsymbol{y}$;
2. $\varphi_r(\boldsymbol{x},\boldsymbol{y})\cdot\varphi_r(\boldsymbol{y},\boldsymbol{x})=1$;
3. 对任意的 $\boldsymbol{z}\in X$,有

$$\varphi_r(\boldsymbol{x},\boldsymbol{y})\cdot\varphi_r(\boldsymbol{y},\boldsymbol{z}) = \theta_r(\boldsymbol{x},\boldsymbol{z});$$

则称 $\varphi_r(\boldsymbol{x},\boldsymbol{y})$是决策个体 DM_r 关于供选方案对$(\boldsymbol{x},\boldsymbol{y})$的个体偏比度.

设 R 是决策群体 G 在 X 上的偏爱关系.若实数$\varphi(\boldsymbol{x},\boldsymbol{y})>0$ 满足:$\varphi(\boldsymbol{x},\boldsymbol{y})\geqslant 1 \Leftrightarrow \boldsymbol{x}R\boldsymbol{y}$,则称 $\varphi(\boldsymbol{x},\boldsymbol{y})$是决策群体 G 关于供选方案对$(\boldsymbol{x},\boldsymbol{y})$的群体偏比度.

较多规则(major rule)　亦称简单多数规则.一种简单和基本的群体决策方法.设 X 是供选方案集,$\boldsymbol{x},\boldsymbol{y}\in X$.群体决策的较多规则是:若决策群体中多数(不少于总人数之半)的决策个体认为 $\boldsymbol{x}$ 严格偏爱于 $\boldsymbol{y}$,则群体即认为 $\boldsymbol{x}$ 严格偏爱于 $\boldsymbol{y}$(参见"偏爱关系").据此,决策群体可以对 X 中的各供选方案做出群体决策的偏爱排序.

α 较多规则(α-major rule)　一种常用的群体决策方法.设 X 是供选方案集,$\boldsymbol{x},\boldsymbol{y}\in X$,决策个体共 $l(\geqslant 2)$人,

$$\alpha \in \left\{\frac{r}{l-r}\middle| r=1,2,\cdots,l-1\right\}\cup\{1\}.$$

群体决策的 α 较多规则是:若决策群体的各决策个体中,认为 $\boldsymbol{x}$ 严格偏爱于 $\boldsymbol{y}$ 者不少于 $\boldsymbol{y}$ 严格偏爱于 $\boldsymbol{x}$ 者的 α 倍,则群体即认为 $\boldsymbol{x}$ 严格偏爱于 $\boldsymbol{y}$(参见"偏爱关系").据此,决策群体可以对 X 中的供选方案做出群体决策的偏爱选优或偏爱排序.当 $\alpha=1$ 时,α 较多规则即较多规则(参见"较多规则").

鲍达数规则(Borda number rule)　一种基本的群体决策方法.设 $X=\{\boldsymbol{x}^1,\boldsymbol{x}^2,\cdots,\boldsymbol{x}^s\}(s\geqslant 3)$是供选方案集,$DM_r$ 是第 r 个决策个体$(r=1,2,\cdots,l;l\geqslant 2)$,$\boldsymbol{x}\in X$.记 $B_r(\boldsymbol{x})$是 DM_r 认为 $\boldsymbol{x}$ 严格偏爱于 $\boldsymbol{y}\in X$ 的 $\boldsymbol{y}$ 的个数,则 $B_r(\boldsymbol{x})$称为 $\boldsymbol{x}$ 关于 DM_r 的鲍达数,而

$$B(\boldsymbol{x}) = \sum_{r=1}^{l} B_r(\boldsymbol{x})$$

称为 $\boldsymbol{x}$ 的鲍达数.设 $\boldsymbol{x}^i,\boldsymbol{x}^j\in X(i,j=1,2,\cdots,s;i\neq j)$,群体决策的鲍达数规则是:若 $B(\boldsymbol{x}^i)\geqslant(>)B(\boldsymbol{x}^j)$,则群体即认为 $\boldsymbol{x}^i$ 偏爱(严格偏爱)于 $\boldsymbol{x}^j$(参见"偏爱关系").据此,决策群体可以对 X 中的各供选方案做出群体决策的偏爱排序.早在 18 世纪后期,鲍达(Borda,J. G.)在研究选举制度中涉及了这一规则.

群体效用函数(group utility function)　描述群体效用的函数.设 X 为供选方案集,$\boldsymbol{x}\in X$,$U_r(\boldsymbol{x})$ $(r=1,2,\cdots,l)$为决策个体 DM_r 对方案 $\boldsymbol{x}$ 的效用,则 $U(\boldsymbol{x})=U_G(U_1(x),U_2(x),\cdots,U_l(x))$称为 X 上群体效用函数.群体效用函数和每个决策个体的效用函数都适合效用的诺伊曼-莫根施特恩公理:若群体中每个决策者都认为某两个方案是无差异的,则群体也认为这两个方案是无差异的.群体效用函数能表示为加法模型:

$$U(\boldsymbol{x}) = \sum_{r=1}^{l}\lambda_r U_r(\boldsymbol{x}),$$

式中的 $\lambda_r(r=1,2,\cdots,l)$是相应的权因子.上式的条件可用另外三个充分条件代替,即:

1. 若群体中所有的个体除 DM_r 外对所有可能的方案都认为无差异,则群体效用函数是 DM_r 的效用函数的正线性变换 $U(\boldsymbol{x})=a+bU_r(\boldsymbol{x})$,即群体的偏爱等价于 DM_r 的偏爱.

2. 若对于所有的 $r=1,2,\cdots,l$,有 $U_r(\boldsymbol{x})=U_0(\boldsymbol{x})$,则群体效用函数 $U(\boldsymbol{x})$必须产生和 $U_0(\boldsymbol{x})$同样的决策.

3. 若群体中所有决策个体除 DM_i 和 DM_j 外,都认为所有可能的供选方案是无差异的,则群体对这些方案所产生的后果的任何展望的偏爱,仅取决于 DM_i 和 DM_j 的偏爱.

把条件 2 修改为条件 3,再联合其他条件,便成为一乘法模型的充分条件.乘法模型为

$$kU(\boldsymbol{x}) + 1 = \prod_{r=1}^{l}[kk_rU_r(\boldsymbol{x}) + 1].$$

群体效用函数 U 和决策个体 $DM_r(r=1,2,\cdots,l)$的效用函数 U_r 的标度值都处在$[0,1]$范围内.k 和 k_r $(r=1,2,\cdots,l)$都是标度常数.

委托效用法(entrust utility method)　一种群体决策方法.是通过委托过程求解群体决策问题的方法.委托过程的步骤如下:

1. 决策群体中决策个体 DM_i 对他委托的小组中各决策个体的效用 $u_j^0(j=1,2,\cdots,l)$设定权 $p_{ij}(j=1,2,\cdots,l)$,并用由此得到的线性组合的效用函数代替其原效用函数 u_i^0,从而

$$u_i^1 = \sum_{j=1}^{l}p_{ij}u_j^0.$$

2. 每个 DM_i 用第一步得到的委托效用函数作

为其效用函数(一般仍采用原来设定的.如有必要,也可以对权作适当的修改).于是,DM_i 的新的委托效用函数为

$$u_i^2 = \sum_{j=1}^{l} p_{ij}u_i^1 \quad (i = 1,2,\cdots,l).$$

3. 把委托过程继续下去,则 DM_i 在第 k 步的委托效用函数为

$$u_i^k = \sum_{j=1}^{l} p_{ij}u_i^{k-1} \quad (i = 1,2,\cdots,l).$$

上式可用矩阵形式表示为

$$\boldsymbol{u}^k = P\boldsymbol{u}^{k-1} = p^k\boldsymbol{u}^0, \tag{1}$$

式中 $\boldsymbol{u}^k=(u_1^k,u_2^k,\cdots,u_l^k)^{\mathrm{T}}$,$P$ 表示元素为 p_{ij} 的 $l\times l$ 矩阵.(1)式能用于两种情形:一是从供选方案集中选择一方案,委托步骤一直进行下去,直到压缩到一个点.也就是说,如果对 X 中每个供选方案 $\boldsymbol{x}$ 计算其效用向量 $\boldsymbol{u}^k(\boldsymbol{x})$,它表示所有决策个体对方案 $\boldsymbol{x}$ 在第 k 步的委托效用.委托过程终止当且仅当在 X 中存在惟一的方案 $\boldsymbol{x}$,其效用 $\boldsymbol{u}^k(\boldsymbol{x})$ 不劣于 X 中的其他任何方案.二是决策群体企图把所有供选方案按群体偏爱排序,从而必须标定群体效用函数.委托效用函数列如果收敛,即向量 $\boldsymbol{u}^k$ 的每个分量都收敛到相同的函数,则委托过程将产生惟一的群体效用函数.这时

$$u = \lim_{k\to\infty} u_i^k \quad (i = 1,2,\cdots,l), \tag{2}$$

式中的 u 是群体效用函数,$u_i^k(i=1,2,\cdots,l)$ 是 DM_i 的第 k 步的委托效用函数.按照效用函数的定义有

$$u = \sum_{i=1}^{l} \lambda_i u_i^0 = \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{u}^0 \tag{3}$$

(其中 $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_l)^{\mathrm{T}}$).由(1)式得

$$u_i^k = \boldsymbol{p}_i^{k^{\mathrm{T}}}\boldsymbol{u}^0, \tag{4}$$

式中的 $\boldsymbol{p}_i^k$ 是矩阵 P 的第 i 行构成的行向量,因此由(2)式、(3)式和(4)式得

$$\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{u}^0 = \lim_{k\to\infty} \boldsymbol{p}_i^{k^{\mathrm{T}}}\boldsymbol{u}^0 \quad (i = 1,2,\cdots,l).$$

由于上式对于任意选择的 $\boldsymbol{u}^0$ 都必须适合,因此有

$$\boldsymbol{\lambda}^{\mathrm{T}} = \lim_{k\to\infty} \boldsymbol{p}_i^k \quad (i = 1,2,\cdots,l). \tag{5}$$

把 P 作为马尔可夫链的一步转移概率矩阵,则第 k 步的 P^k 是这个链的 k 步转移概率矩阵.由(5)式,这个链有如下的性质:

$$\lim_{k\to\infty}\boldsymbol{p}_1^k = \lim_{k\to\infty}\boldsymbol{p}_2^k = \cdots = \lim_{k\to\infty}\boldsymbol{p}_l^k = \boldsymbol{\lambda}^{\mathrm{T}}.$$

即要求 P^k 中的元素 $p_{ij}(k)$ 对于一切 i,j,存在不依赖于 i 的极限,即

$$\lim_{k\to\infty} p_{ij}(k) = \lambda_j,$$

$p_{ij}(k)$ 是此链的 k 步转移概率.称此性质为遍历性.对于一有限马尔可夫链,如果存在正整数 N,使

$$p_{ij}(N) > 0 \quad (i,j = 1,2,\cdots,l), \tag{6}$$

则此链是遍历的,且(3)式中 $(\lambda_1,\lambda_2,\cdots,\lambda_l)$ 是方程组

$$\lambda_j = \sum_{i=1}^{l} \lambda_i p_{ij} \quad (j = 1,2,\cdots,l)$$

满足条件

$$\lambda_j > 0, \quad \sum_{i=1}^{l} \lambda_j = 1$$

的惟一解.使用此结果可以求得一组权 $(\lambda_1,\lambda_2,\cdots,\lambda_l)$.这里需要满足充分条件(6)式,只须群体中的每个决策个体设定其委托小组的权 $p_{ij}(i,j=1,2,\cdots,l)$ 的值能满足(6)式即可.除此法外,还可以运用纳什(Nash,J.F.)的谈判模型求群体决策问题的解.

群体效用函数法(group utility function method)　一种群体决策方法.是由群体价值判断求群体效用函数的方法.其步骤如下:

1. 把决策群体中的各决策个体分配成各个子群 $\{G_1,G_2,\cdots,G_m\}$.这种分配是基于决策个体对属性的专长.对于属性 $y_i(i=1,2,\cdots,m)$ 有专门知识的那些决策个体负责估计群体的边际效用函数 $u_{Di}(y_i)$.

2. 决策个体被分配到各子群以后,需要作两种类型的群体的价值判断.一是子群 G_i 估计属性 y_i 的群体边际效用 u_{Di}.二是估计 $u_D(y_1,y_2,\cdots,y_m)=H(u_{D1}(y_1),u_{D2}(y_2),\cdots,u_{Dm}(y_m))$,$H$ 表示算子,$u_D(y_1,y_2,\cdots,y_m)$ 是群体效用函数,$u_{Di}(y_i)(i=1,2,\cdots,m)$ 是属性 y_i 的群体边际效用函数中算子 H 的参数.算子 H 一般取相加或相乘的形式.

值得指出的是,这里要做出的价值判断不是人与人之间的效用比较,而是估计属性间的置换.两种类型价值判断的量化均类似于效用函数构造法的量化.主要区别是使用某决策规则把群体中每个决策个体的反应集结为群体的反应.若提供给群体一系列的抽奖,则群体效用函数能被估计.群体效用函数估计的关键是一集结过程,把群体中每个决策个体的价值判断集结为群体的价值判断,只要群体中全体决策个体同意选择一个大家都愿意接受的决策规则即可.

NGT 法(nominal group technique)　一种逐步形成群体意见的群体决策方法.该方法的思路是:对一群体决策问题,发动群体中各决策个体提意见,然后按一定程序集结决策个体的意见,以做出群体的判断.此法适合 5～9 个决策个体的较小规模的群体决策问题.方法步骤如下:

1. 每个决策个体在安静的环境下写出自己的意见.

2. 组织者不分先后地听取并记录这些意见.

3. 集体逐步讨论这些意见,弄清楚它们的意义.

4. 对归纳意见所形成的条目的重要性作初步投票.

5. 讨论初步投票.

6. 最终投票.

德尔菲法(Delphi method) 亦称专家调查法. 一种用规定程序对专家进行调查的群体决策方法. 此法要尽量精确地反映出专家们的主观估计能力，是经济调查法中的一种比较科学的方法. 德尔菲(Delphi)是古希腊的一个城市的名字，是预言家们的活动场所. 德尔菲法的参加人员包括：一是决策者，二是专业干部(这两部分人构成德尔菲法的组织者)，三是反应者(他们是德尔菲法征求意见的对象，通常是各方面的专家，是制订决策的群体成员). 方法分为以下 9 步：

1. 提出问题.

2. 选择并确定群体中的决策个体(反应者).

3. 制订第一个咨询表，并散发给各决策个体.

4. 收集第一个咨询表，并进行分析.

5. 制订第二个咨询表，并散发给各决策个体.

6. 收集第二个咨询表，并对数据进行统计处理.

7. 制订第三个咨询表，并散发给各决策个体.

8. 收集第三个咨询表，并对新的数据进行统计处理.

9. 准备最后的报告.

自 1964 年美国兰德(R&D)公司的专家们提出德尔菲法后，又派生了几种修正的德尔菲法.

群体多目标决策问题(multicriteria decision making problem of group) 具有多于 1 个目标的群体决策问题. 是多目标决策问题和群体决策问题的一种综合(参见“多目标决策问题”、“群体决策问题”). 极小化形式的群体多目标规划问题的数学模型(GVP)是

$$\{V-\min_{x\in X} \boldsymbol{f}^r(\boldsymbol{x}), r=1,2,\cdots,l\},$$

其中 X 是供选方案集，$\boldsymbol{f}^r: X\to R^{m_r}$ 是第 $r(r=1,2,\cdots,l)$ 个决策个体的多目标函数，记号“$V-\min$”的意义参见“多目标规划”. 对于此类模型，已给出多种解的概念，如群体效用有效解、综合效用最优解、群体一致有效解，以及各种联合有效解等. 同时，已提出求解此类问题的一些效用函数方法和交互规划算法(参见“评价函数方法”和“交互规划算法”).

群体效用有效解(group utility efficient solution) 群体多目标决策问题的解. 对于群体多目标规划问题(GVP)(参见“群体多目标决策问题”)，设 $u_r: R^{m_r}\to R(r=1,2,\cdots,l)$ 是第 r 个决策个体的效用函数(增函数)，则 $(u_1(\boldsymbol{f}^1(\boldsymbol{x})), u_2(\boldsymbol{f}^2(\boldsymbol{x})),\cdots, u_l(\boldsymbol{f}^l(\boldsymbol{x})))^{\mathrm{T}}$ 是 $\boldsymbol{x}$ 的多目标函数. 若 $\tilde{\boldsymbol{x}}\in X$，并且不存在 $\boldsymbol{x}\in X$，使得 $u_r(\boldsymbol{f}^r(\boldsymbol{x}))\leqslant u_r(\boldsymbol{f}^r(\tilde{\boldsymbol{x}}))$ $(r=1,2,\cdots,l)$，并且其中至少有一个是严格不等式，则称 $\tilde{\boldsymbol{x}}$ 是群体多目标规划问题(GVP)的群体效用有效解. 若 $\tilde{\boldsymbol{x}}\in X$，并且不存在 $\boldsymbol{x}\in X$，使得 $u_r(\boldsymbol{f}^r(\boldsymbol{x}))<u_r(\boldsymbol{f}^r(\tilde{\boldsymbol{x}}))$ $(r=1,2,\cdots,l)$，则称 $\tilde{\boldsymbol{x}}$ 是(GVP)的群体效用弱有效解.

综合效用最优解(synthetic utility optimal solution) 群体多目标决策问题的一种解. 对于群体多目标规划问题(GVP)(参见“群体多目标决策问题”)，设 $u_r: R^{m_r}\to R(r=1,2,\cdots,l)$ 是第 r 个决策个体的效用函数(增函数)，$u: R^l\to R$ 是群体效用函数，

$$\boldsymbol{u}(\boldsymbol{x})=\boldsymbol{u}(u_1(\boldsymbol{f}^1(\boldsymbol{x})), u_2(\boldsymbol{f}^2(\boldsymbol{x})),\cdots, u_l(\boldsymbol{f}^l(\boldsymbol{x})))^{\mathrm{T}}.$$

若 $\boldsymbol{x}^*\in X$，并且对任意的 $\boldsymbol{x}\in X$，有 $\boldsymbol{u}(\boldsymbol{x}^*)\leqslant\boldsymbol{u}(\boldsymbol{x})$，则称 $\boldsymbol{x}^*$ 是群体多目标规划问题(GVP)的综合效用最优解.

群体一致有效解(group uniform efficient solution) 群体多目标决策问题的解. 设

$$(P)_r: V-\min_{x\in X} \boldsymbol{f}^r(\boldsymbol{x})$$

是群体多目标规划问题(GVP)中对应于第 $r(r=1,2,\cdots,l)$ 个决策个体的多目标规划问题(参见“群体多目标决策问题”). 若 $\tilde{\boldsymbol{x}}\in X$ 同时是各 $(P)_r$ 的(弱)有效解，则称 $\tilde{\boldsymbol{x}}$ 是群体多目标规划问题(GVP)的群体一致(弱)有效解(参见“有效解”和“弱有效解”).

较多联合有效解(major joint efficient solution) 群体多目标决策问题的一类解. 设 $\boldsymbol{x}\in X$，令

$$\mu^r(\boldsymbol{x})=\begin{cases}1 & (\boldsymbol{x}\text{ 是}(P)_r\text{ 的有效解}),\\ 0 & (\boldsymbol{x}\text{ 不是}(P)_r\text{ 的有效解})\end{cases}$$

(记号 $(P)_r$ 参见“群体一致有效解”)，$i, N\in\{1,2,\cdots,l\}$. 若 $\tilde{\boldsymbol{x}}\in X$，并且

$$\sum_{r=1}^{l}\mu^r(\tilde{\boldsymbol{x}})=i,$$

则称 $\tilde{\boldsymbol{x}}$ 是群体多目标规划问题(GVP)的 i 联合有效解. 若 $\tilde{\boldsymbol{x}}\in X$，并且

$$\sum_{r=1}^{l}\mu^r(\tilde{\boldsymbol{x}})\geqslant N,$$

则称 $\tilde{\boldsymbol{x}}$ 是问题(GVP)的 N 联合有效解(参见“群体多目标决策问题”).

单目标序贯决策问题(sequential decision problem of an objective) 一类决策问题. 是只有一个目标函数的序贯决策问题. 它分为确定性单目标序贯决策问题和随机性单目标序贯决策问题两种：

1. 确定性单目标序贯决策问题. 选择 $u_i(i=1,2,\cdots,n-1)$，使

$$\begin{cases}\min\left\{\sum_{i=0}^{n-1}v_i(x_i,u_i)+v_n(x_n)\right\},\\ \text{s.t. } x_{i+1}=T_i(x_i,u_i)\\ \qquad (i=0,1,2,\cdots,n-1),\end{cases}\tag{1}$$

x_i 和 u_i 分别属于有限集 X_i 和 U_i. 令从第 i 个状态 x_i 开始到最终状态 x_n，则(1)式改写为

$$v_i(x_i, u_i, \cdots, x_{n-1}, u_{n-1}, x_n)$$
$$(i = 0, 1, \cdots, n-1). \tag{2}$$

设(2)式满足

$$v_i(x_i, u_i, \cdots, x_n)$$
$$= \Phi_i[x_i, u_i, v_{i+1}(x_{i+1}, u_{i+1}, \cdots, x_n)] \tag{3}$$

的递推关系，对于给定的 $x_i \in X_i$，用 $\pi_i(x_i)$ 表示序列 $\{u_i, \cdots, u_{n-1}\}$，则 v_i 为 x_i 和 π_i 的函数，记为 $J_i(x_i, \pi_i)$，$\pi_i(i=0,1,\cdots,n-1)$ 称为策略. 策略 $\pi_i(x_i)$ 的集合记为 $N_i(x_i)$，(3)式可改写为 $J_i(x_i, \pi_i) = \Phi_i[x_i, u_i, J_{i+1}(x_{i+1}, \pi_{i+1})]$，式中的 x_i，u_i 和 x_{i+1} 也具有(1)中约束条件的递推关系.

2. 随机性单目标序贯决策问题. 随机性序贯决策问题和确定性序贯决策问题主要不同点是：前者的状态是随机转移的，当决策给定时，它形成随机过程，即马尔可夫过程. 离散的马尔可夫过程，称为马尔可夫链，相应的决策问题，称为马尔可夫决策问题. 马尔可夫决策问题的最优决策是在某段时间内的期望损失最小. 从 $t+1$ 个周期到过程终止的期望损失为 $J_{t+1}(x^j, \pi_{t+1})$，式中的 x^j 是第 $t+1$ 个周期的状态，π_{t+1} 是从第 $t+1$ 个周期到过程终止的决策规则 δ 的序列 $\{\delta_{t+1}, \delta_{t+2}, \cdots\}$，称之为策略. 决策规则 δ_t 为一函数，它把第 t 个周期的每个状态 $x_t \in X$ 映射为决策 $u_t \in U$. δ_t 的集合记为 Δ. 从第 t 个周期到过程终止的期望损失为

$$J_t(x^i, \pi_t) = \sum_{j=1}^{m} p_{ij}^k (r_{ij}^k + J_{t+1}(x^j, \pi_{t+1})), \tag{4}$$

式中 $x^i = x_t$ 是第 t 个周期的状态，p_{ij}^k 表示由状态 x^i 转移到 $x^j = x_{t+1}$ 的一步转移概率，损失 r_{ij}^k 不随周期变化，$\pi_t = (\delta_t, \pi_{t+1})$，$\delta_t$ 是第 t 周期的决策规则. 若用 $v(x^i, u^k)$ 表示状态 x^i 和决策 $u^k = u_t$ 时，下一周期中所遭受的期望损失即

$$v(x^i, u^k) = \sum_{j=1}^{n} p_{ij}^k r_{ij}^k,$$

则(4)式可改写为

$$J_t(x^i, \pi_t)$$
$$= v(x^i, u^k) + \sum_{j=1}^{n} p_{ij}^k J_{t+1}(x^j, \pi_{t+1}). \tag{5}$$

此式是解马尔可夫决策问题的基本方程. 若用符号 i 表示状态 $x_i(i=1,2,\cdots,m)$，$J_i(t)$ 表示 $J_t(x^i)$，v_i 表示 $v(x^i)$，$\boldsymbol{J}(t) = (J_1(t), \cdots, J_m(t))^{\mathrm{T}}$，$\boldsymbol{v} = (v_1, \cdots, v_m)^{\mathrm{T}}$，略去已知的策略 π_{i+1} 和决策规则 δ_t，且把第 t 个和第 $t+1$ 个周期的排列次序颠倒过来，则(5)式改写为

$$\boldsymbol{J}(t+1) = \boldsymbol{v} + P\boldsymbol{J}(t),$$

式中的 P 是 $m \times m$ 矩阵，其第 i 行、第 j 列的元素是 p_{ij}，第 i 个分量是

$$J_i(t+1) = v_i + \sum_{j=1}^{m} p_{ij} J_j(t)$$
$$(i=1,2,\cdots,m; t=0,1,2,\cdots).$$

多目标序贯决策问题(decision problem of multiobjective sequential)　一类决策问题. 是有多个目标函数的序贯决策问题. 多目标序贯决策问题可分为以下几类：

1. 序贯决策过程为一确定的、有限周期的过程. 不考虑相邻两周期状态的转移，而代之以状态集的简单划分，即把状态的一非空任意集 $\mathscr{X}$ 划分为一有序的集合 $(X_1, X_2, \cdots, X_n)$，$X_1, X_2, \cdots, X_n$ 是非空的不相交的子集，X_t 表示过程在第 t 个周期的可能状态的集. 此外，不采用实值的效用函数，而代之以策略的优先关系，策略的优先关系为全预序(即弱序). 这种分析方法称为优序动态规划.

2. 序贯决策的过程是确定的、有限周期的，状态从一周期按给定的规则转移到下一周期. 多目标的损失函数为一向量函数. 策略的优先关系是偏预序. 寻求此序贯决策问题的帕雷托有效解(参见“帕雷托解”).

3. 序贯决策过程是确定的、无限周期的，状态从一周期按给定的规则转移到下一周期，不采用实值的效用函数，而代之以策略的优先关系. 策略的优先关系为全预序. 这种分析方法称为顺序动态规划. 此法可推广到随机过程.

4. 序贯决策过程为无限周期的马尔可夫决策过程，状态从一周期按给定的转移概率转移到下一周期. 多目标的报酬函数为一折扣的向量函数. 策略的优先关系为偏预序. 寻求此序贯决策问题的帕雷托有效解.

以上四类问题按其策略的优先关系可区分为全预序和偏序. 若是全预序，定义一序数效用函数，寻求最佳调和解. 但优先关系为偏预序时，则借助凸锥寻找有效解(参见“锥有效解”). 这四类问题又可划分为有限周期的和无限周期的. 求解多目标序贯决策问题的常用方法为优序动态规划方法和顺序动态规划方法.

图 与 网 络

图与网络理论(graph and network theory)　运筹学与组合数学及代数等的交叉学科. 作为运筹学的分支，它主要研究图论和网络理论中有关最优化的问题. 图论是研究一组离散对象间二元关系的数学分支. 这种关系结构可以抽象成图的概念，并直观地用几何表示和研究. 在数学科学中，图论属于组合学，而与拓扑、代数等学科关系密切. 图论的起源可以追溯到 1736 年欧拉(Euler, L.)关于哥尼斯堡

七桥问题(参见“哥尼斯堡七桥问题”)的研究,而作为一门独立学科却是20世纪50年代形成的.图论在物理、化学、电子学、通信科学、计算机科学、经济学、语言学、心理学等领域均有重要应用.网络理论研究网络的结构和最优化问题.其系统理论源于电路分析,与图论和数学规划相结合而形成.网络最优化作为运筹学的分支,近几十年来发展迅速,在科学管理、控制论、信息论、通信理论,以及计算机技术中得到有效应用.

图论与网络理论中与运筹学紧密联系的一些最优化问题包括:最小树问题、最短有向路问题、分派问题、中国邮递员问题和旅行售货员问题,以及网络流问题等.这些重要而基本的内容有着广阔而深刻的实际背景,在电子技术、生产分配系统、交通运输系统、通信系统、财贸系统、军事后勤系统、管理网络系统,以及电网络系统中有着广泛的应用.

哥尼斯堡七桥问题(problem of seven bridges of Königsberg) 一个古老的图论问题.在哥尼斯堡城(现名加里宁格勒),有一条普雷格尔(Pregel)

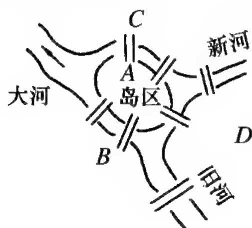


图1

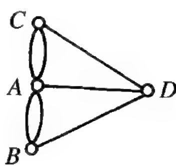


图2

河,河中有个岛,岛上有七座桥,如图1所示.当地居民热衷于这样一个问题:一位散步者能否从一点出发,走过这七座桥,且每座桥只走一次,最后回到出发点.1736年,欧拉(Euler,L.)把这个问题归结为如图2所示的一笔画问题,即能否从一点开始,不重复地一笔画出此图.答案是否定的.

图(graph) 一类组合构形.一个图就是集合V和E的一个二元组$G(V,E)$,其中V称为顶点集,E是V的一部分或全部顶点对的集合,称为边集.当考虑E中元素视为V中无序对时,称G为无向图(图1);而考虑V中元素的有序对(不必相异)时,称G为有向图(图2).V的元素称为顶点,E的元素称为边(无向图)或弧(有向图).当图G的一条边e是G的一对顶点(u,v)时,称e连结u和v;顶点u和v称为e的端点,一条边的端点称为与这条边关联,与同一条边关联的两个顶点称为邻接的(或相邻的),与同一个顶点关联的两条边也称为邻接的(或相邻的).

无向图(undirected graph) 见“图”.

有向图(digraph or directed graph) 见“图”.有向图的基础图是指把该图中的弧用边来代替而得

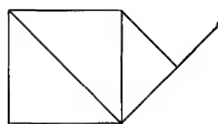


图1 无向图

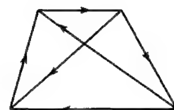


图2 有向图

到的无向图.有向图常记为D,其中的弧的集合记为A.

网络(network) 一类带权的图.设G是一个图(有向图),若对G的每条边(弧)赋予一个实数,并称为这条边(弧)的权,则G连同它的边(弧)上的权w称为网络(有向网络),记为$G=(V,E,w)$($D=(V,A,w)$).一个图(有向图)可以看做是所有边(弧)的权都为1的网络(有向网络).上面的符号见“图”、“有向图”.

树(tree) 一类特殊的图.一个无圈的连通图,称为树.在树中,任意两顶点之间必有一条且仅有一条链.在树中,去掉任一条边,则树成为不连通图.在树中不相邻的两个顶点间添上一条边,构成一个圈.设T为n个顶点的一棵树,则T的边数为$n-1$.

支撑子图(spanning subgraph) 亦称生成子图.一类图.由一个图的全部顶点及连结这些顶点的部分边构成的图称为原图的支撑子图.给定图$G=(V,E)$,若$V'\subseteq V$,$E'=\{uv\in E\mid u,v\in V'\}$,则称图$G'=(V',E')$是由$V'$生成的子图,记为$G'=G[V']$.

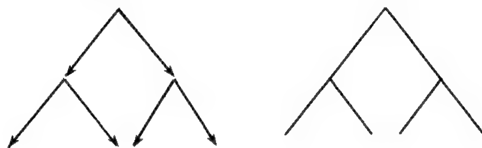
支撑树(spanning tree) 亦称生成树.一类特殊的树.图G的一个支撑子图T是一棵树,则称T为G的支撑树.G有支撑树当且仅当G是连通的.

最小树(minimum spanning tree) 一类支撑树.指权最小的支撑树.给定网络$G=(V,E,w)$,设$T=(V',E')$为G的一棵支撑树,令

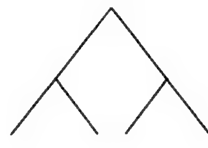
$$w(T)=\sum_{e\in E'}w(e),$$

则$w(T)$称为T的权.G中权最小的支撑树称为G的最小树.

树形图(arborescence) 一类有向树.如果有向



树形图D



D的基础图

图D(参见“有向图”)满足下列条件,则称D为树形图:

1. D的基础图是树.

2. D中仅有一个顶点的λ度为0,其余顶点的λ度都是1.

λ度为0的顶点称为这个树形图的根.如图所

示，一个顶点处的 λ 度数，即指向该顶点的箭头数.

最小树形图(minimum arborescence)　一类有向树. 设 $H=(V,A',w)$ 是有向网络 $D=(V,A,w)$ 的一个支撑树形图，则称

$$w(H)=\sum_{a\in A'}w(a)$$

为 H 的权. 而权最小的支撑树形图称为最小树形图. 此条目中的符号参见"网络"和"支撑树".

最小树问题(minimal tree problem)　网络最优化问题之一(参见"网络最优化问题"). 如何从网络的支撑树中求出最小树的问题，称为最小树问题. 求解最小树问题常用破圈法和贪婪算法.

最小树形图算法(minimum arborescence algorithm)　一种求图上有向树的方法. 它的基本思想是：任给有向网络 G，先对每一点选一条权最小的入弧，再从这 n 条弧中选 $n-1$ 条权较小的弧，然后观察由这 $n-1$ 条弧所构成的 G 的支撑子图是否包含有向圈. 若含有有向圈，则通过收缩有向圈得到一个新的网络. 这网络的弧数与点数总比原来的少. 对这新网络再继续上面的方法，于是便得到一个网络的递减序列：$G_0,G_1,\cdots,G_k$. 这时，要么 G_k 不具有支撑树形图，从而 G 也不具有支撑树形图；要么 G_k 有一个最小树形图. 因此，可按 $G_k,\cdots,G_1,G_0$ 的顺序逐步展开，便得到一个最小树形图的序列 $H_k,\cdots,H_1,H_0$. H_0 是 G 的最小树形图.

克鲁斯卡尔算法(Kruskal algorithm)　亦称贪婪算法. 寻求最小树的最早的一类算法. 算法之一是：每步从未选的边中选一条最小权的边，使之与已选边不构成圈；若遇到两条或几条边的权相等时，可从中任选一条边. 算法之二是：任取一圈，从圈上去掉一条最大权的边，在余下的圈中，重复这个步骤，直到无圈时为止，即可求出最小树；若遇到两条或几条边的权相等时，可任意丢去其中一条边.

贪婪算法(greedy algorithm)　即"克鲁斯卡尔算法".

避圈法(avoid cycle method)　求图上支撑树的一种方法. 在图中任意取一条边 e_1，找一条不与 e_1 构成圈的边 e_2. 然后，再找一条与 $\{e_1,e_2\}$ 不构成圈的边 e_3，继续找下去，直到这个过程不能进行时为止. 这时所得到的图就是一棵支撑树.

破圈法(tear cycle method)　求图上支撑树的一种方法. 在一个无环连通图中取一个圈，从圈中丢去任一边，对余下的图重复这个步骤，直到无圈为止，于是得到一棵支撑树.

最短路(shortest path)　权最小的路. 在赋权有向图中，从始点 v_1 到终点 v_p 总权最小的一条有向路，称为 v_1 至 v_p 间的最短有向路，简称最短路.

最短路问题(shortest path problem)　一类组合最优化问题. 给定一个有向网络 $D=(V,A,w)$，设 P 为 D 中的一条有向路，令

$$w(P)=\sum_{a\in P}w(a),$$

则 $w(P)$ 称为 P 的权. 最短路问题就是寻求有向网络中自某一指定点 i 到另一指定点 j 间的权最小的有向路. 最短路问题是网络流中最基本的一个课题(参见"网络流"). 一大类组合最优化问题都可以化为求最短有向路，或者用最短有向路的算法作为其子程序. 例如，通信网络中的最可靠路问题、最大容量路问题，统筹方法中求关键路线以及有名的行囊问题，都可以化为求最短有向路问题. 另外，某些加工顺序问题、中国邮递员问题，都要用最短有向路算法作为其子程序. 因此，最短有向路的用途已远超出了其直观的意义. 最短有向路的算法与有向网络中权的正负情形有很大关系. 在具有负有向圈(有向圈的权为负值)的有向网络中，至今还没有求最短路的有效算法. 求最短有向路的算法比较著名的有帝克斯彻算法、代换法等；求所有点对间最短有向路问题有弗罗特方法；对庞大而稀疏的有向网络问题有分解方法；还有求前 m 条最短路的算法.

最短路算法(shortest path algorithm)　亦称帝克斯彻算法. 一种求最短路问题的多项式算法(参见"最短路"和"最短路问题"). 有两种方法.

1. 只适合于非负权的情况. 先从始点 v_1 开始，给每一个顶点记一个数，称为标号. 标号分 T 标号和 P 标号两种. T 标号表示从始点 v_1 到这点的最短路权的上界，称为临时标号；P 标号表示从 v_1 到该点的最短路权，称为固定标号. 已得到 P 标号的点不再改变，凡是没有标上 P 标号的点，标上 T 标号. 算法的每一步把某一点的 T 标号改变为 P 标号. 对于具有 p 个点的问题，最多经过 $p-1$ 步，就可以得到从始点到每一点的最短路. 在求出最短路权之后，再采用"反向追踪"的办法，可以求出最短路线.

2. 适合于有负权的情况. 设从任一点 v_i 到任一点 v_j 都有一条弧. 若不含点 v_k 到点 v_l 的弧，则添加一条 (v_k,v_l)，并规定其权 $w(v_k,v_l)=+\infty$. 设要求的是从点 v_1 到一点 $v_j(j=2,3,\cdots,p)$ 的最短路，其权为 l_{1j}，简记为 l_j. 从 v_1 到 v_j 的最短路总是先从 v_1 沿着一条路到某一点 v_i，再沿 (v_i,v_j) 到 v_j. 而从 v_1 到 v_i 这条路必然是从 v_1 到 v_i 的最短路，即路的权为 l_i. 所以 l_i 必须满足方程 $l_j=\min\{l_i+w_{ij}\}$. 为了求此方程的解：$l_1,l_2,\cdots,l_p$，可令 $l_j^{(1)}=w_{1j}(j=1,2,\cdots,p)$，对 $t=2,3,\cdots$，求 $l_j^{(t)}=\min\{l_i^{(t-1)}+w_{ij}\}$. 当用此递推公式进行到某步 k，且对所有的 $j=1,2,\cdots,p$，有 $l_j^{(k)}=l_j^{(k-1)}$ 时，则 $l_1^{(k)},l_2^{(k)},\cdots,l_p^{(k)}$ 即为从 v_1 到各点的最短路的权. 递推公式中的 $l_j^{(t)}$ 是从 v_1 到 v_j 的至多含有 $t-1$ 个中间点的最短路权. 在一个具有 p

个点的图中，若不含有总权小于零的圈（称为负圈），则从始点 v_1 到任一点的最短路最多含有 $p-2$ 个中间点. 所以，当 D 中不含负圈时，这种算法最多经过 $p-1$ 次迭代必收敛，即对所有的 $j=1,2,\cdots,p$ 有 $l_j^{(k)}=l_j^{(k-1)}$. 若经过 $p-1$ 次迭代，有某个 j，使得 $l_j^{(p)}\neq l_j^{(p-1)}$，此时图中有负圈，最短路权是无下界的. 为了加快收敛速度，可用

$$l_j^{(1)}=w_{1j}\ (j=1,2,\cdots,p),$$
$$l_j^{(t)}=\min\{\min_{i<j}(l_i^{(t)}+w_{ij}),\min_{i\geqslant j}(l_i^{(t-1)}+w_{ij})\}$$
$$(j=1,2,\cdots,p,\ t=2,3,\cdots).$$

在求出最短路权之后，再采用"反向追踪"的办法，可以求出最短路线. 最短路算法是由帝克斯彻(Dijkstra，E. W.)于 1959 年首先提出的. 该算法是目前较好的算法.

帝克斯彻算法(Dijkstra algorithm)　即"最短路算法".

弗罗特方法(Floyd method)　在有向网络上求所有点对之间最短有向路的一种方法. 已有求自一点到其余各点最短有向路的计算方法，只要将这个方法使用 n 次，就可求出所有点对之间的最短有向路，但这样做计算量太大. 弗罗特方法是比此更好的算法. 设 U_{ij}^m 表示自点 i 到点 j 且不经过点 $m,m+1,\cdots,n$（除去点 i 和点 j）的最短有向路的长度，则自点 i 到点 j 不经过 $m+1,m+2,\cdots,n$（除去点 i 和点 j）的最短有向路长度 $U_{ij}^{m+1}=\min\{U_{ij}^m,U_{im}^m+U_{mj}^m\}$. 当 $m=n$ 时，U_{ij}^{m+1} 就等于网络中自点 i 到点 j 的最短有向路的长度. 具体算法如下：

1. 置 $m=1$，且 $U_{ij}^m=w_{ij}$ 而 $w_{ii}=0$，$r_{ij}^m=j(j=1,2,\cdots,n)$.

2. 对一切 i,j，计算 $U_{ij}^{m+1}=\min\{U_{ij}^m,U_{im}^m+U_{mj}^m\}$. 若 $U_{ij}^m>U_{im}^m+U_{mj}^m$，则 $r_{ij}^{m+1}=r_{ij}^m$；若 $U_{ij}^m\leqslant U_{im}^m+U_{mj}^m$，则 $r_{ij}^{m+1}=r_{ij}^m$.

3. 若 $m=n$，则步骤终止. 根据 (r_{ij}^m) 内的记录寻找出所有 i 和 j 之间的最短有向路. 若 $m<n$，则 $m=m+1$，返回步骤 2.

弗罗特算法的计算量为 $O(n^3)$.

欧拉迹(Eulerian trail)　图论术语. 指图的一条通过图中每条边恰好一次的途径. 图的一条闭欧拉迹称为该图的欧拉环游. 含有欧拉环游的图称为欧拉图. 欧拉(Euler，L.)于 1736 年证明了：当且仅当非空连通图的每一顶点都是偶次时，它才是欧拉图.

欧拉环游(Euler tour)　见"欧拉迹".

欧拉图(Eulerian graph)　见"欧拉迹".

哈密顿圈(Hamiltonian cycle)　图论术语. 指图中包含所有顶点的圈. 一个图若含有哈密顿圈，则称为哈密顿图.

哈密顿图(Hamiltonian graph)　见"哈密顿圈".

中国邮递员问题(Chinese postman problem)　一类组合最优化问题. 问题是：一个邮递员递送邮件，从邮局出发，要走完他负责投递的全部街道，完成任务后回到邮局，应该按怎样的路线走，使所走的路程最短. 中国邮递员问题和旅行售货员问题是组合最优化中两个著名问题. 若把邮递员所管辖的街道的平面图看做是一个网络 $G=(V,E,w)$（参见"网络"），其中 V 中的点表示路口，E 中的边表示街道，边权 w 表示边所对应的街道长，则中国邮递员问题便转化为在 G 中求一条通过所有边的权最小的回路. 若 G 的所有点均为偶点，则 G 包含经过每条边恰好一次的回路（称为欧拉环游，参见"欧拉迹"），这时，这条欧拉环游便是所要求的最短路线. 如果 G 上含有奇点，可通过增加一些重复边，得到一个新的欧拉型网络 G^*，使 G^* 中不含奇点. 问题转化为在 G 中找一组权最小的边，使当这些边变成双重时，G 的每一点的次数都变成偶数（参见"奇偶点图上作业法"和"厄得蒙斯算法"）. 中国邮递员问题是中国学者管梅谷首先研究的，因而得名.

福楼里算法(Fleury algorithm)　在图上求欧拉环游的一种方法. 中国邮递员问题实际上是在具有非负权的连通网络 G 中找出一条最小权的通过所有边的闭途径. 当 G 是欧拉图时，问题转化为在 G 中确定一条欧拉环游（参见"欧拉迹"）. 福楼里算法的基本原则是：每到一点，沿该点的关联边中未走过的一条走，只有当没有其他选择时，才选未走过边所导出的子图中的割边. 步骤如下：

1. 任意选取一个顶点 v_0，置 $w_0=v_0$.

2. 假设迹 $w_i=v_0e_1v_1\cdots e_iv_i$ 已经选定，那么按下述方法从 $E\backslash\{e_1,e_2,\cdots,e_{i-1},e_i\}$ 中选取边 e_{i+1}：

1) e_{i+1} 和 e_i 相关联.

2) 除非没有别的边可选择，否则 e_{i+1} 不是 $G_i=G-\{e_1,e_2,\cdots,e_i\}$ 的割边.

3. 当步骤 2 不能再执行时，算法终止.

奇偶点图作业法(graphical method based on an odd-even-point approach)　求最优邮递路线的一种方法. 其特点是：对有奇点的街道，方法是求邮递员递送邮件的路线为最短. 设在某条路线中的边 $[v_i,v_j]$ 上重复走了几次，在 v_i,v_j 之间增加几条边，令每条边的权和原来的权相等，并把新增加的边称为重复边. 这样，就把原问题变为：在一个有奇点的图中，要求增加一些重复边，使新图不含奇点，并且重复边的总权为最小. 把使新图不含奇点而增加的重复边，简称可行（重复边）方案. 使总权最小的可行方案称为最优方案. 如此把原问题作如下转变：

1. 第一个可行方案的确定方法. 若图中有奇点，

则把它配成对,每一对奇点之间必有一条链,把这条链的所有边作为重复边加到图中去,新图中必无奇点.便给出了第一个可行方案.

2. 调整可行方案,使重复边总长度下降.当边$[v_i, v_j]$上有两条或两条以上的重复边时,从中去掉偶数条,得到一个总长度较小的方案.于是,有:

1) 在最优方案中,图的每条边上最多有一条重复边.

2) 在最优方案中,图中每个圈上的重复边的总权不大于该圈总权的一半.

3. 判断最优方案的标准.一个最优方案一定是满足上述1)和2)的可行方案.反之,一个可行方案若满足上述1)和2),则这个可行方案一定是最优方案.

根据判断标准,对给定的可行方案,检查它是否满足上述条件1)和2).若满足,所得方案即为最优方案;若不满足,则对方案进行调整,直至上述条件1)和2)均得到满足时为止.

厄得蒙斯算法(Edmonds algorithm) 求解中国邮递员问题的一种方法.主要思想是:当网络G不是欧拉图时,可通过边的双重运算来消灭奇点,从而得到新的欧拉型网络G^*.这样,中国邮递员问题便化为在G中找一组权和最小的边,使得这些边双重时,G的每一点都是偶点.此法是厄得蒙斯(Edmonds, J.)于1973年提出.其算法步骤如下:

1. 在网络G中辨认奇数次点,若不存在,则转向步骤4.

2. 计算所有奇数次点之间的最短路.

3. 构造完全网络G',G'的点集合由G的奇点组成,G'的边权$w'_{ij}=l-l_{ij}$,其中l是一个大数,l_{ij}表示G中奇数次点i和j之间的最短路的长度,用求最大权对集算法,求出G'的最大权对集M.

4. G'中M的每一条边都对应G中两个奇点间的一条最短路,取G中对应于G'中M的所有最短路的边的和集E',E'即为所求.

旅行售货员问题(traveling salesman problem) 一类组合最优化问题.设有一个售货员从城市1出发,到城市2,3,…,n去推销货物,最后回到城市1.假定任意两个城市i,j间的距离d_{ij}($d_{ij}=d_{ji}$)是已知的,问他应沿着什么样的路线走,才能使走过的路线最短?容易看出,中国邮递员问题要求走遍所有"线",而后者要求走遍所有"点".旅行售货员问题就是在一个完全网络中,找出一个具有最小权的哈密顿圈.寻求旅行售货员问题的有效算法似乎是没有希望的,它属于NP完全类.一个可行的办法是首先求一个哈密顿圈,然后适当修改,以得到具较小权的另一个哈密顿圈.旅行售货员问题有着明显的实际意义,除售货员之外,邮局里负责到各个信箱取信的邮递员,以及去各个分局送邮件的汽车等都会类似地遇到这个问题.还有一些问题表面上似乎与之无关,而实质上却可以归结为旅行售货员问题求解,如计算机线路问题、无中间存储的工件加工问题等.

对集(matching) 亦称匹配.一种特殊的子图.给定一个图$G=(V,E)$,设M是E的一个子集.若M中任意两条边均不相邻,则称M为图G的一个对集(或匹配).若对集M的某条边与顶点v关联,则称M饱和顶点v,并且称v是M饱和的,否则称v是M非饱和的.若G的每个顶点均为M饱和,则称M为G的完美对集.若不存在另一对集M',使$|M'|>|M|$,则称M为G的最大基数对集,简称最大对集.完美对集一定是最大对集.反之不成立.

匹配(matching) 即"对集".

最大对集(maximum matching) 见"对集".图G的对集M是最大对集当且仅当G不包含M增长路(参见"M增长路").

完美对集(perfect matching) 见"对集".

M交错路(M-alternating path) 图论术语.是伴随一个对集的路(参见"对集").设M是图$G=(V,E)$的一个对集,G的一条M交错路是指其边在M和$E\setminus M$中交错出现的路.

M增长路(M-augmenting path) 一种M交错路.图G的M增长路是指起点和终点都是M非饱和的一条M交错路(参见"对集").亦称M可扩路.

匈牙利法(Hungarin method) 一种求最大对集的算法.是二部图$(X,Y;E)$上求最大对集的一个有效算法(参见"对集").计算量为$O(|X|^2|Y|)$.其基本思想为:从任意一个对集M开始,按以下步骤进行:

1. 若M饱和X的每个顶点,则停止;否则,设u是X中的M非饱和顶点,置$S=\{u\}$及$T=\varnothing$.

2. 若$N(S)=T$,则停止;否则,设$y\in N(S)\setminus T$.

3. 若y是M饱和的,设$yz\in M$,用$S\cup\{z\}$代替S,$T\cup\{y\}$代替T,并转到步骤2;否则,设P是M增广$(u-y)$路,在P上将M的边去掉并加进非M的边代替M,转到步骤1.

最大对集问题(maximum matching problem) 一类组合最优化问题.指在一个给定图上找一个最大对集(最大基数对集)的问题(参见"对集").二部图(偶图)的最大对集问题可用匈牙利法求解,也可转化为网络的最大流问题求解.1965年,厄得蒙斯(Edmonds, J.)对匈牙利法作了修改,得到了一般图最大对集的一个有效算法.

最大-最小对集问题(max-min matching problem) 一类组合最优化问题.指在给定一个二部网络$G=(X,Y;E,w)$上,求G的一个最小边权达到最

大的最大对集(参见“对集”).存在求这种对集的一种有效算法.

最大权对集问题(weighted matching problem) 一种组合最优化问题.指在给定网络 $G=(V,E,w)$ 上,求 G 的一个具有最大权的对集问题(参见“对集”).最大权对集问题存在有效算法.

人员分派问题(personnel assignment problem) 亦称第一类分派问题.一种特殊的运输问题.某公司准备分派 n 个工人 $x_1,x_2,\cdots,x_n$ 做 n 件工作 $y_1,y_2,\cdots,y_n$,已知这些工人中每个人都胜任一件或几件工作,试问能否把所有工人都分派做一件他所能胜任的工作?这就是人员分派问题.构造一个二部图 $G=(X,Y,E)$,其中 $X=\{x_1,x_2,\cdots,x_n\}$,$Y=\{y_1,y_2,\cdots,y_n\}$,并且 $\{x_i,y_j\}\in E$ 当且仅当 x_i 能胜任 y_j 工作.于是,人员分派问题转化为求 G 的完美对集问题(参见“对集”).它可用求二部图最大对集的匈牙利法得以解决.

最优分派问题(optimal assignment problem) 亦称第二类分派问题.一种特殊的运输问题.问题是:某工厂准备分派 n 个工人 $x_1,x_2,\cdots,x_n$ 做 n 件工作 $y_1,y_2,\cdots,y_n$.已知工人 x_i 做工作 y_j 的效益为 $w_{ij}\geqslant 0$,现求一种分派方案,使得总的效益达到最大.考察一个完全二部图 (X,Y),$X=\{x_1,x_2,\cdots,x_n\}$,$Y=\{y_1,y_2,\cdots,y_n\}$.边 x_iy_j 有权 $w_{ij}=w(x_i,y_j)$,表示工人 x_i 做工作 y_j 时的效益.最优分派问题显然等价于在一个赋权完全二部图中寻找一个最大权的对集,称为最优对集(参见“对集”).求解此问题,可以用求最大权对集的算法,也可用更为简便的库恩-曼克尔斯算法(参见“库恩-曼克尔斯算法”).

最优对集(optimal matching) 见“最优分派问题”.

最小基数最优对集(minimum cardinality optimal matching) 一种最优对集.指权大于零的边尽可能少的最优对集(参见“对集”和“最优分派问题”).在最优分派问题中,若每个工人作任何一种工作所创造的价值都大于零,则最优对集一定是完美对集,即每一个工人都被分派做一种工作.否则,最优对集不一定是完美对集,即可能不需要 n 个工人一样能创造出最大价值.在这种情况下,最小基数最优对集就是要求以最少的人去创造最大总价值.

最小级差最优对集(minimum difference optimal matching) 一种最优对集.指最大边权与最小边权之差尽可能小的最优对集(参见“最优分派问题”).在最优分派问题中,若计划人员不仅希望创造的总价值最大,而且希望每一个人创造的价值尽可能平均,则对应于找一个最小级差最优对集.

库恩-曼克尔斯算法(Kuhn-Munkres algorithm) 一种求解最优分派问题的方法.若顶点集 $X\cup Y$ 上的实值函数 l 适合下述条件:对所有的 $x\in X$ 以及 $y\in Y$,均有 $l(x)+l(y)\geqslant w(x,y)$,则把这个函数定义为该二部图的一个可行顶点标号(实数 $l(v)$ 称为顶点 v 的标号).不管边的权是什么,总存在可行点标号.如下述定义的函数 l 就是这样的可行点标号:

$$\begin{cases} l(x)=\max\limits_{y\in Y}\{w(x,y)\} & (x\in X),\\ l(y)=0 & (y\in Y).\end{cases}$$

令 $E_l=\{xy\in E\mid l(x)+l(y)=w(x,y)\}$,具有边集 E_l 的 G 的生成子图称为对应于可行点标号 l 的相等子图,用 G_l 表示.库恩-曼克尔斯算法建立在如下定理的基础上:设 l 是 G 的可行顶点标号,若 G_l 包含完美对集 M^*,则 M^* 是 G 的最优对集(参见“对集”和“最优分派问题”).

可平面图(planar graph) 一类特殊的图.若一个图能画在平面上使得它的边仅在端点相交,则称这个图为可嵌入平面的,或称为可平面图.可平面图 G 的这样一种画法,称为 G 的一个平面嵌入.G 的一个平面嵌入 $\widetilde{G}$ 本身可看做是同构于 G 的图,所以有时将可平面图的平面嵌入就称为平图.图 G 可嵌入平面当且仅当它可嵌入球面.一个图是平面图当且仅当它不含 K_5 或 K_{33}的剖分图.

平图(plane graph) 见“可平面图”.

桥(bridge) 一种组合构形.设 H 是图 G 的一个给定子图,在 $E(G)\backslash E(H)$ 上定义关系“$\sim$”如下:$e_1\sim e_2$ 当且仅当存在一条途径 W,使得:

1. W 的第一条边和最后一条边分别是 e_1 和 e_2.

2. W 与 H 是内部不相交的(即 W 的内部顶点没有一个是 H 的顶点).

容易验证“$\sim$”是$E(G)\backslash E(H)$上的等价关系.由关系“$\sim$”的等价类导出的 $G-E(H)$的子图称为 G 中 H 的桥.

平面性算法(planarity algorithm) 一种求平面嵌入的算法.指判断一个给定的图是否是可平面图,并且如果它是可平面图,则要找出它的平面嵌入来.设 H 是图 G 的一个可平面子图,$\widetilde{H}$ 是 H 在这个平面中的一个嵌入,若 G 是可平面图,且存在 G 的一个平面嵌入 $\widetilde{G}$,使得 $\widetilde{H}\subseteq\widetilde{G}$,则称 $\widetilde{H}$ 是 G 容许的.若 B 是 H(在 G 中)的任意一座桥,且 B 对于 H 的接触点都包含在 $\widetilde{H}$ 的面 f 的周界上,则称 B 在 $\widetilde{H}$ 的面 f 内是可画的.用 $F(B,\widetilde{H})$ 表示在 $\widetilde{H}$ 中桥 B 是可画出的那些面的集.平面性算法是基于如下定理:若 $\widetilde{H}$ 是容许的,则对于 H 的每座桥 B,$F(B,\widetilde{H})\neq\varnothing$.步骤如下:

1. 设 G_1 是 G 的一个圈,找出 G_1 的一个平面嵌入 $\widetilde{G}_1$.置 $i:=1$.

2. 若 $E(G)\backslash E(G_i)=\varnothing$，则停止；否则，确定 G 中 G_i 的所有桥；对于每座这样的桥 B，找出集 $F(B,\widetilde{G}_i)$.

3. 若存在一座桥 B，使得 $F(B,\widetilde{G}_i)=\varnothing$，则停止. 根据定理，$G$ 是非平面图. 若存在一座桥 B，使得 $|F(B,\widetilde{G}_i)|=1$，则令 $\{f\}=F(B,\widetilde{G}_i)$，否则令 B 是任一座桥，且 f 是任一使得 $f\in F(B,\widetilde{G}_i)$ 的面.

4. 选择一条连结 B 对于 G_i 的两个接触顶点的路 $P_i\subseteq B$，置 $G_{i+1}=G_i\cup P_i$，并把 P_i 画在 $\widetilde{G}_i$ 的面 f 内，得到 G_{i+1} 的一个平面嵌入 $\widetilde{G}_{i+1}$，令 $i+1:=i$，转入步骤 2.

网络流(flow in network) 网络中弧集上的一种函数. 对于给定的网络 $D=(V,A,c)$，定义弧集 A 上的一个实值函数 f 满足以下条件，则称 f 为网络 D 上的一个流，并称 $f(a)$ 为弧 a 上的流量：

1. 对所有 $a\in A, 0\leqslant f(a)\leqslant c(a)$.

2. 对所有 $v\in V$，$f^-(v)=f^+(v)$，其中 $f^+(v)=\sum f(a)$，a 取遍所有以 v 为始点的弧；$f^-(v)=\sum f(a)$，a 取遍所有以 v 为终点的弧.

网络流理论研究的基本问题有：最短路问题、最小树问题、最小费用流问题等. 它在信息论、控制论、工程技术、交通运输、社会集团结构、化学键结构、语言结构等方面有着广泛的应用.

可行流(feasible flow) 一种满足特定条件的流. 即满足容量限制条件和平衡条件的流.

1. 容量限制条件：对每一弧 $(v_i,v_j)\in A$，

$$0\leqslant f_{ij}\leqslant c_{ij}.$$

2. 平衡条件：对于中间点，流出量＝流入量，即对每个 $i(i\neq s,t)$ 有

$$\sum_{(v_i,v_j)\in A} f_{ij}-\sum_{(v_j,v_i)\in A} f_{ji}=0,$$

对于发点 v_s，记

$$\sum_{(v_s,v_j)\in A} f_{sj}-\sum_{(v_j,v_s)\in A} f_{js}=V(f)$$

($V(f)$ 称为这个可行流的流量，即发点的净输出量或收点的净输入量)，则对于收点 v_t，有

$$\sum_{(v_t,v_j)\in A} f_{tj}-\sum_{(v_j,v_t)\in A} f_{jt}=-V(f).$$

零流(zero flow) 一种特殊的流. 设 f 为网络上的流，若所有弧的流量均为零，则称此流为零流(参见“网络流”). 零流的流量为 $v(f)=0$.

容量(capacity) 一种权函数. 在有向图 $D=(V,A)$ 的弧集 A 上定义非负函数 c，称函数 c 为容量函数，它在弧 $a\in A$ 上的值称为 a 的容量.

饱和弧(saturated arc) 一种带有特定性质的弧. 设 $f=\{f_{ij}\}$ 为网络上给定的可行流，则使 $f_{ij}=c_{ij}$ 的弧 (v_i,v_j) 称为 f 的饱和弧；称使 $f_{ij}<c_{ij}$ 的弧为非饱和弧；称使 $f_{ij}=0$ 的弧为零流弧；称使 $f_{ij}>0$ 的弧为非零流弧. 设 μ 是一条链或路，若链中弧的方向与链的方向一致，称为前向弧，它们的全体记为 μ^+. 若弧的方向与链的方向相反，称为后向弧，记它们的全体为 μ^-.

增广路(augmenting path) 权函数满足特定性质的路. 从发点 v_s 到收点 v_t 的一条路 μ，若 μ 满足：在弧 $(v_i,v_j)\in\mu^+$ 上，$0\leqslant f_{ij}<c_{ij}$，即 μ^+ 中每一弧是非饱和弧(参见“饱和弧”). 在弧 $(v_i,v_j)\in\mu^-$ 上，$0<f_{ij}\leqslant c_{ij}$，即 μ^- 中每一弧是非零流弧，称 μ 为关于可行流 f 的一类增广路.

截集(cut set) 有向图上弧的一种特定的子集. 给定有向网络 $D=(V,A,C)$. 若 V 被剖分为两个非空集合 $V_1,\overline{V}_1$. 使 $v_s\in V_1, v_t\in\overline{V}_1$，称弧集 $(V_1,\overline{V}_1)$ 是分离 v_s 和 v_t 的截集. 若把某一截集的弧从网络 $D=(V,A,C)$ 中丢去，则从 v_s 到 v_t 便不存在有向路. 因此，截集是从 v_s 到 v_t 的必经之路.

截量(cut magnitude) 网络的基本概念之一. 指截集中所有弧的容量之和. 记为 $C(V_1,\overline{V}_1)$，即

$$C(V_1,\overline{V}_1)=\sum_{(v_i,v_j)\in(V_1,\overline{V}_1)} c_{ij}.$$

任何一个可行流的流量 $V(f)$ 都不超过任一截集的截量，即 $V(f)\leqslant C(V_1,\overline{V}_1)$(参见“网络流”).

最大流问题(maximum flow problem) 一种组合最优化问题. 网络流理论研究的一个基本问题是求网络中一个可行流 f，使其流量 $V(f)$ 达到最大，这种流 f 称为最大流(参见“网络流”)，这个问题称为(网络)最大流问题. 最大流问题是一个特殊的线性规划问题，一般用标号法寻求最大流比用求线性规划问题的一般方法要方便得多.

最小费用流问题(minimum-cost flow problem) 一种组合最优化问题. 对给定网络 $D=(V,A,C)$，在弧集 A 上定义非负函数 b，对 $a\in A$，$b(a)$ 表示弧 a 上单位流量的费用，称 b 为费用函数. 最小费用流问题，即对一给定流量 val，求一可行流 f，使其满足：

1. 流量 $v(f)=\text{val}$.

2. $\sum_{a\in A} b(a)f(a)$ 最小.

最小费用流问题是网络流理论研究的一个重要问题.

最大流的最小截集准则(maximum flow and minimum cut set) 用最小截集判定最大流的准则(参见“最大流问题”). 若对于一个可行流 f^*，网络 $D=(V,A,C)$ 中有一个截集 $(V_1^*,\overline{V}_1^*)$，使

$$V(f^*)=C(V_1^*,\overline{V}_1^*),$$

则 f^* 必是最大流，且 $(V_1^*,\overline{V}_1^*)$ 必是 D 的所有截集中容量最小的一个，即最小截集. 可行流 f^* 是最大流当且仅当不存在关于 f^* 的增广路. 寻求最大流的

一个方法是：若给了一个可行流 f，只要判断 D 中有无关于 f 的增广路. 若有，改进 f，得到一个流量增大的新的可行流. 若无，则得到最大流.

最小费用最大流（minimum-cost maximum flow） 一种达到某类极值的流. 指一个最大流 f，它使流的总输送费用

$$b(f)=\sum_{(v_i,v_j)\in A} b_{ij}f_{ij}$$

取极小值（参见“最大流的最小截集准则”）. 先考察当沿着一条关于可行流 f 的增广路 μ，以 $\theta=1$ 调整 f，得到新的可行流 f'（参见“可行流”和“增广路”）. 此时，

$$\begin{aligned}&b(f')-b(f)\\&=\Big[\sum_{\mu^+} b_{ij}(f'_{ij}-f_{ij})-\sum_{\mu^-} b_{ij}(f'_{ij}-f_{ij})\Big]\\&=\sum_{\mu^+} b_{ij}-\sum_{\mu^-} b_{ij}.\end{aligned}$$

称

$$\sum_{\mu^+} b_{ij}-\sum_{\mu^-} b_{ij}$$

为这条增广路 μ 的“费用”. 若 f 是流量为 $V(f)$ 的所有可行流中费用最小者，而 μ 是关于 f 的所有增广路中费用最小的增广路，则沿 μ 调整 f，得到的可行流 f' 就是流量为 $V(f')$ 的所有可行流中的最小费用流. 当 f' 是最大流时，它就是所要求的最小费用最大流. 设已知 f 是流量 $V(f)$ 的最小费用流，为了寻求关于 f 的最小费用增广路，构造一个赋权有向图 $w(f)$，它的顶点是原网络 D 的顶点，而把 D 中的每一条弧 (v_i,v_j) 变成两个相反方向的弧 (v_i,v_j) 和 (v_j,v_i). 定义 $w(f)$ 中弧的权 w_{ij} 为

$$w_{ij}=\begin{cases}b_{ij} & (f_{ij}<c_{ij}),\\+\infty & (f_{ij}=c_{ij}),\end{cases}$$

$$w_{ij}=\begin{cases}-b_{ij} & (f_{ij}>0),\\+\infty & (f_{ij}=0)\end{cases}$$

（长度为 $+\infty$ 的弧可以从 $w(f)$ 中略去）. 于是，在网络 D 中寻求关于 f 的最小费用增广路，等价于在赋权有向图 $w(f)$ 中寻求从 v_s 到 v_t 的最短路. 其算法如下：开始取 $f^{(0)}=0$. 若在第 $k-1$ 步得到最小费用流 $f^{(k-1)}$，则构造赋权有向图 $w(f^{(k-1)})$，在 $w(f^{(k-1)})$ 中，寻求从 v_s 到 v_t 的最短路. 若不存在最短路（即最短路权是 $-\infty$），则 $f^{(k-1)}$ 就是最小费用最大流；若存在最短路，则在原网络 D 中得到相应的增广路 μ，在增广路 μ 上对 $f^{(k-1)}$ 进行调整，调整量为

$$\theta=\min\{\min_{\mu^+}(c_{ij}-f_{ij}^{(k-1)}),\min_{\mu^-}(f_{ij}^{(k-1)})\}.$$

令

$$f_{ij}^{(k)}=\begin{cases}f_{ij}^{(k-1)}+\theta & ((v_i,v_j)\in\mu^+),\\f_{ij}^{(k-1)}-\theta & ((v_i,v_j)\in\mu^-),\\f_{ij}^{(k-1)} & ((v_i,v_j)\notin\mu),\end{cases}$$

得到新的可行流 $f^{(k)}$. 再对 $f^{(k)}$ 重复上述步骤.

最小费用最大流问题（minimum-cost maximum flow problem） 一种运输问题. 对给定网络 $N=(V,A,C,b)$（其中 b 为费用函数），最小费用最大流问题就是求一最小费用流，使其满足流量条件 $V(f)=V^*$，其中 V^* 为 N 的最大流量（参见“最大流问题”）. 最小费用最大流问题是最小费用流问题的一个特例（参见“最小费用流问题”）.

寻求最大流的标号法（labelling method of solution to maximum flow） 以标号过程为特征寻求最大流的方法. 即从一个可行流 f 出发，经过标号过程与调整过程，最终得到一个最大流的方法（参见“可行流”和“最大流问题”）.

1. 标号过程. 网络中的点，或是标号点（分为已检查和未检查两种），或是未标号点. 标号点的标号包含两部分：第一个标号表明它的标号是从哪一点得到的，以便找出增广链；第二个标号是为确定增广链的调整量 θ 用的. 标号过程开始时，总先给发点 v_s 标上 $(0,+\infty)$，这时 v_s 是标号而未检查的点，其余都是未标号点. 一般地，取一个标号而未检查的点 v_i，对一切未标号点 v_j：

1) 若在弧 (v_i,v_j) 上有 $f_{ij}<c_{ij}$，则给 v_j 标号 $(v_i,l(v_j))$，这里 $l(v_j)=\min[l(v_i),c_{ij}-f_{ij}]$，此时点 v_j 成为标号而未检查的点.

2) 若在弧 (v_j,v_i) 上有 $f_{ji}>0$，则给 v_j 标号 $(-v_i,l(v_j))$，这里 $l(v_j)=\min[l(v_i),f_{ji}]$，此时 v_j 成为标号而未检查的点.

于是，v_i 成为标号而已检查过的点. 重复上述步骤，一旦收点 v_t 被标号，表明得到一条从 v_s 到 v_t 的增广链 μ，转入调整过程. 若所有标号都已检查过，而标号过程进行不下去时，则算法结束，此时的可行流就是最大流.

2. 调整过程. 按收点 v_t 及其他点的第一个标号，利用“反向追踪”的办法，找出增广链 μ. 设 v_t 的第一个标号为 v_k（或 $-v_k$），则弧 (v_k,v_t)（或 (v_t,v_k)）是 μ 上的弧. 接着检查 v_k 的第一个标号，若为 v_i（或 $-v_i$），则找出 (v_i,v_k)（或 (v_k,v_i)）. 再检查 v_i 的第一个标号，依此进行，直到 v_s 为止. 这时，被找出的弧就构成了增广链 μ. 令调整量 θ 为 $l(v_t)$，即 v_t 的第二个标号. 令

$$f'_{ij}=\begin{cases}f_{ij}+\theta & ((v_i,v_j)\in\mu^+),\\f_{ij}-\theta & ((v_i,v_j)\in\mu^-),\\f_{ij} & ((v_i,v_j)\notin\mu).\end{cases}$$

去掉所有的标号，对新的可行流 $f'=\{f'_{ij}\}$ 重新进入标号过程. 用标号法找增广链以求最大流的结果，同时得到一个最小截集.

统　筹　法

统筹法(program evaluation and review technique)　一类网络上的最优化方法.指在制定计划和进行管理时,统筹兼顾、合理安排的方法.统筹法以数学为工具,通过网络的形式,进行数量分析,实施控制和调整,增强计划的科学性、有效性,达到省时、省力、减少消耗的目的.统筹法现已发展为运筹学的一类重要的实用方法.统筹法是 20 世纪 50 年代发展起来的,在美国称为计划评估技术(Program Evaluation and Review Technique,简称 PERT),也有人称为关键路线法(Critical Path Method,简称 CPM).1964 年,数学家华罗庚在中国推广了这种科学方法,并根据其主要特点——统筹安排,定名为统筹法.统筹法在工农业生产的计划管理、科学实验和军事领域的组织指挥、后勤保障等方面都有广泛的应用.

关键路线法(critical path method)　即"统筹法".

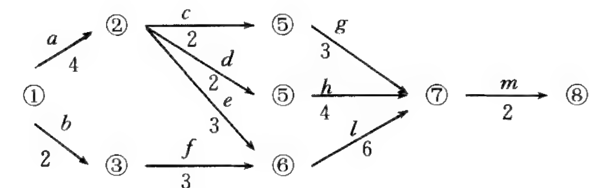
网络方法(network method)　一种统筹法.绘制网络图的规划及计算相关参数(其中主要是时间参数)的方法.

网络计划(net program)　一种以网络形式表示的计划.是以网络图表示并用网络方法编制的计划.

工序(work)　亦称工作.统筹法术语.指为完成某一项任务而进行的各项相对独立的活动.用一支箭(→)表示一道工序,把表示各道工序的各支箭,按照工程的工艺顺序,从左至右,逻辑地排列起来,就可以画出一个表示一系列工序的图.

结点(node)　统筹图的基本元素.是相邻工序的分界点(参见"工序").结点用圆圈表示,连结箭尾的结点表示工序的开始,连结箭头的结点表示工序的完成.

事项(item)　统筹法术语.是相邻工序在时间上的分界点(参见"工序").表示工序的开工或完工,用注有编号的结点表示.在统筹图中,用一支箭及开工事项、完工事项表示一道确定的工序,如图所示.



图中的事项①表示工程开始和 a,b 工序开工,事项 ② 表示 a 工序完工及 c,d,e 三道工序开工……,事项 ⑧ 表示工程的最后一道工序 m 完工,即整个工程完工.把与代表工序的一支箭的箭尾相接的开工事项称为该道工序的箭尾事项;把与箭头相接的完工事项称为该道工序的箭头事项.工序的箭尾事项与箭头事项称为工序的相关事项.

统筹图(network chart)　亦称网络图或箭头图.一种有向图(参见"图").指由工序、事项及标有完成各道工序所需时间等参数所构成的有向图(参见"工序"和"事项").用箭线表示工序,结点表示事项的统筹图,称为箭线式统筹图.用结点表示工序,箭线表示事项的统筹图,称为结点式统筹图.

网络图(network chart)　即"统筹图".

箭头图(arrow chart)　即"统筹图".

绘制统筹图的规则(rule of design network chart)　一种制作逻辑图方法的规则.绘制统筹图一般应遵循如下 9 条规则:

1.方向、时序与编号.按照工艺流程的顺序,工序从左向右排列.一支箭和它的相关事项只能表示一道工序.一道工序的相关事项只能表示一道工序的开工和完工.在绘制时,要如实反映工序之间的衔接关系及时序,并按照时序对事项编号.

2.紧前工序与紧后工序.若有工序 b 和 c,都需要在工序 a 完工后才能开工,如图 1 所示.工序 b 和 c 的紧前工序是 a,工序 a 的紧后工序是 b,c .

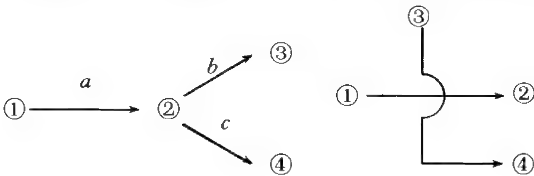


图 1　　　　图 2

3.缺口与回路.统筹图中从始至终沿任何路都可达到终点,不能有缺口.统筹图中也不能有回路,即工序流程不能出现循环,否则,表示工程永远不能完工.

4.虚工序.它是虚设的,仅用来表示相邻工序之间的衔接,并用虚线表示,不需要时间、费用和资源.

5.平行作业.在编制计划时,为加快工程速度,有些工序可以同时进行,即采用平行作业的方式.为便于确定关键路线,在绘制几道工序平行作业完工后转入下一道工序时,选择在平行作业的几道工序中工序时间最长的一道工序,将它直接通过事项与下一道工序衔接,而其他工序则通过虚工序与下一道工序衔接.用虚工序表示工序的平行作业就可以避免两道或几道工序都具有相同的事项编号,可防止绘图及编制计算机程序的混乱.

6.交叉作业.对需要较长时间完成的相邻几道工序,只要条件允许,不必等到紧前工序全部完工后

再转入后一道工序，而是分批分期地将紧前工序完成的部分任务转入下一道工序，这种方式称为交叉作业. 它可以缩短工程周期.

7. 始点与终点. 一张统筹图，只能用一个始点和一个终点表示工程的开工和完工. 在工程开工时几道平行作业的工序中，如果有两道工序和几道工序同时是某一道工序的紧前工序时，在始点后设置虚工序，可使统筹图清晰明了；在工程完工的最后几道平行作业的工序中，如果有两道或几道工序同时是某一道工序的紧后工序时，应在终点前设置虚工序.

8. 简化与合并. 将工序综合程度较低的统筹图，合并为工序综合程度较高的统筹图，从而可使工序数目减少，称为统筹图的简化，但要注意简化前后的时间参数要一致. 把若干个局部统筹图合并为一个统筹图，称为统筹图的合并.

9. 箭线的画法及统筹图的布局. 箭线用水平线或具有一段水平线的折线画出，不画或少画交叉线，在无法避免交叉线时，交叉线可以直接通过，如图 2 所示. 当箭线过于密集时，用一条“母线”与各箭线连结. 统筹图的布局要突出重点，关键路线要布置在图的中心位置，把联系紧密的工序布置在一起. 有时也可绘制附有时间进度和完成各道工序的工作单位的统筹图.

结点式统筹图(node type network chart) 一种统筹图. 是以结点表示工序、箭线表示工序之间联系的统筹图. 结点用圆圈表示，圆圈内标注工序代号和该道工序所需的时间. 在实际工作中，应用较为广泛的是箭线式统筹图(参见“统筹图”).

关键路线(critical path) 亦称主要矛盾线. 完成各道工序需要时间最长的路线. 关键路线用粗线或双线，或红线标出. 在时间进度上，关键路线是完成工程计划的关键. 关键路线上的工序称为关键工序. 编制统筹计划主要是在一个庞大的统筹图中找出关键路线，在各道关键工序上挖潜革新，以达到缩短工期、降低工程费用、提高工程质量和合理利用资源的目的. 关键路线是相对的，也可根据实际进行修改.

时间参数(time parameter) 统筹图的基本元素. 指统筹图中的某一延续时间(即某一段时间)或某一时刻. 由于绘制统筹图时，要找出关键路线，故使用最多的是时间参数.

延续时间(continue time) 一种特定的时间段. 有以下三种：

1. 工序时间 $t(i,j)$. 为完成某一工序所需的时间. 确定它有两种方法：

1）根据工时定额或劳动定额资料确定工序时间. 若无这些资料，可根据已有的这方面或同类工序的时间消耗统计资料，用分析对比法确定工序时间.

2）不具备 1)，且工序时间较长，可用三时估计法确定工序时间，即

$$t=\frac{a+4m+b}{6},$$

其中 a 表示最乐观时间，是顺利完成工序的最短时间；b 表示最保守时间，是不顺利完成工序的最长时间；m 表示最可能时间，是正常地完成工序的时间. 这里，t 是一个加权平均值，其方差为

$$\sigma^2=\left(\frac{b-a}{6}\right)^2,$$

均方差为

$$\sigma=\frac{b-a}{6}.$$

在一项工程中，工程完工时间等于各道关键工序的平均工序时间之和. 有 S 道关键工序的工程的完工时间服从以

$$T_{\mathrm{E}}=\sum_{i=1}^{S}\frac{a_i+4m_i+b_i}{6}$$

为均值，以

$$\sigma=\sqrt{\sum_{i=1}^{S}\left(\frac{b_i-a_i}{6}\right)^2}$$

为均方差的正态分布. 具有一定概率值的工程完工时间为 $T_{\mathrm{K}}=T_{\mathrm{E}}+\sigma\lambda$ 或 $\lambda=(T_{\mathrm{K}}-T_{\mathrm{E}})/\sigma$，式中 T_{K} 表示规定的工程完工时间或目标时间，T_{E} 表示工程最早完工时间，即关键路线上各工序平均工序时间的总和，λ 表示 σ 的系数.

2. 工序总时差 $R(i,j)$. 在不影响工程最早完工时间的前提下，工序的完工期可以推迟的时间. 也称为工序总的允许延误时间. 工序总时差等于该工序的最迟必须开工时间减去最早可能开工时间，或者是工序的最迟必须完工时间减去工序最早可能完工时间，即 $R(i,j)=t_{\mathrm{LS}}(i,j)-t_{\mathrm{ES}}(i,j)$ 或 $R(i,j)=t_{\mathrm{LF}}(i,j)-t_{\mathrm{EF}}(i,j)$. 关键路线上的各道关键工序，它们的总时差为零. 因此，在统筹图中，以总时差为零的工序所形成的路，来确定关键路线.

3. 工序单时差 $r(i,j)$. 在不影响紧后工序的最早可能开工时间的前提下，工序最早可能完工时间可以推迟的时间.

时差反映了统筹图计划中各道工序及各条路线上工序时间的平衡.

统筹图中的时间(time in network chart) 一些特指的时刻和时段. 统筹图中的某一时间一般特指如下 7 种时间：

1. 工序最早可能开工时间 $t_{\mathrm{ES}}(i,j)$. 它等于它的紧前工序的最早可能完工时间，也称为工序的最早开工期.

2. 工序最早可能完工时间 $t_{\mathrm{EF}}(i,j)$. 它等于最早可能开工时间加上工序时间，即

$t_{EF}(i,j)=t_{ES}(i,j)+t(i,j)$.

3. 工程最早完工时间 T_E. 在统筹图中的各道关键工序,都按照最早可能开工时间开工,这时的工程完工时间称为工程最早完工时间,或称为工程最早完工期. 一般情况下,把工程最早完工时间作为工程完工时间.

4. 工序最迟必须开工时间 $t_{LS}(i,j)$. 在不影响工程最早完工时间的前提下,对于非关键路线上的各道工序可以允许在一定范围内推迟一定的时间开工,它亦称为最迟必须开工期.

5. 工序最迟必须完工时间 $t_{LF}(i,j)$. 又称为最迟必须完工期. 它等于最迟必须开工时间加上工序时间,即 $t_{LF}(i,j)=t_{LS}(i,j)+t(i,j)$.

6. 事项最早时间 $t_E(i)$. 是一个事项最早可能开始的时间. 它等于从始点到本事项的最长路线上各道工序的工序时间之和. 始点事项的最早时间为零,即$t_E(1)=0$.

7. 事项最迟时间 $t_L(j)$. 一个事项若晚于某一时间,就会推迟工程最早完工时间,称此时间为事项最迟时间. 终点事项的最迟时间等于工程最早完工时间,即 $t_L(n)=T_E$,其中 n 为终点事项的编号.

计算时间参数(compute time parameter) 确定特别的时刻或时段的数值. 对时间参数的计算一般有如下几种:

1. 确定工序时间.

2. 计算事项的最早时间. 自左至右逐个事项向前计算,直至算到终点为止. 用公式表示为:$t_E(1)=0,t_E(j)=\max\limits_{i}[t_E(i)+t(i,j)](j=1,2,\cdots,n)$,式中 $t_E(1)$表示始点事项,n 表示终点事项的编号.

3. 计算事项的最迟时间. 自右至左逐个事项逆推计算,直至算到始点为止. 用公式表示为:$t_L(n)=t_E(n),t_L(i)=\min\limits_{j}[t_L(j)-t(i,j)]$.

4. 工序的最早可能开工时间. 有两种计算方法:

1) $t_{ES}(i,j)=t_E(i)$.

2) $t_{ES}(j,k)=\max[t_{ES}(i,j)+t(i,j)]$

$$(j=2,\cdots,n-1),$$

式中 j,k 表示工序的箭尾事项与箭头事项的编号,i 为紧前的一道工序或几道工序的箭尾事项的编号.

5. 工序的最迟必须开工时间. 有两种算法:

1) $t_{LS}(i,j)=t_L(j)-t(i,j)$.

2) $t_{LS}(i,j)=\min[t_{LS}(j,k)-t(i,j)]$.

6. 计算工序的单时差和总时差.

除上述方法外,还有计算时间参数的表格法和计算时间参数的矩阵法.

统筹图的优化(optimization for network chart) 统筹法的重要概念. 指对统筹图的流程、资源和时间等方面的优化. 根据统筹图的主要参数,对统筹图进行分析、调整、合理地安排工作流程,合理地分配资源,缩短完成任务的总时间,降低总费用,寻求最优计划方案,称此过程为统筹图的优化.

丁渭挖沟(Dingwei digs trench) 中国古代体现统筹思想的一个典型例子. 公元 1015 年,北宋汴京城(今河南省开封市)皇宫失火,将皇宫烧成废墟. 宋真宗命丁渭主持修复皇宫的工程. 丁渭首先把皇宫前的大街开挖成一条大沟,取土烧砖、烧瓦;再把京城附近的汴水引进大沟内,使运送建筑材料的船只直抵宫前;竣工后又把废弃的砖瓦和废土填入沟内,然后修复原来的大街. 这一统筹施工安排,可谓"举一役而三得".

模 型 论

模型(model) 现实客观事物的一种表示和体现. 它可以是文字、图表、公式,也可以是计算机程序或其他实体模型,具有以下三个特点:

1. 是现实世界一部分的模仿或抽象.

2. 由那些与分析问题有关的因素构成.

3. 体现了有关因素之间的关系.

模型与现实客观事物相比,其优点是简单、经济、便于操作和试验、运转周期短. 通过对模型的试验,可以对实际问题做出客观的分析.

按模型的形式可分为形象模型和抽象模型两大类;按其中参量的性质可分为确定性模型和随机性模型两大类. 形象模型与实际物体基本相似,如原子模型、飞机模型、儿童玩具、相片等. 抽象模型是用符号、图表等来描述客观事物. 它又分为模拟模型、数学模型和概念模型三类. 确定性模型是指其中的参量是确定量的模型,如线性规划模型、非线性规划模型、网络模型、投入产出模型、确定性存贮模型和排序模型等. 随机性模型是指其中的参量含有随机变量的模型,如随机性存贮模型、排队模型、对策模型、决策模型、预测模型,以及维修、更新、可靠性模型等.

建立模型的方法有以下几种:

1. 直接分析法. 当问题比较简单或比较明显时,按问题的性质和范围直接建立模型.

2. 模拟法. 有些模型的结构性质虽已清楚,但用数量描述它以及求解却很困难. 如果有另一种系统及其结构性质与之相同,构造出的模型也类似,处理时较简单,这时可用后一模型模拟前一模型,对后一模型进行试验和求解.

3. 利用数据分析. 对系统功能的数据进行分析,搞清楚系统的结构模型,用回归分析方法建立模型,并进一步分析一些因素的作用.

4. 概率统计分析法. 利用概率统计方法分析问

题和建立模型.

5. 试验分析法. 现有数据的分析不能确定个别因素对系统工作指标的影响时,有时需要在系统上作局部试验,以搞清哪些是本质变量,以及它对指标的影响.

6. 想定法. 是人为地、科学地设想一些情况,使用已知方法构造出模型,推出结果,然后根据这些结果进行讨论,分析其是否可行. 如果不可行,再重新想定,或者不断修改,经过多次迭代,直至达到所需的结果为止.

建立模型的步骤如下:

1. 明确目标.

2. 对系统进行周密调查,找出主要因素,确定主要变量.

3. 找出各种关系.

4. 明确系统的约束条件.

5. 规定符号、代号.

6. 根据有关学科的知识,用数学符号、数学公式表达所有关系.

7. 有时需要对模型进行修正或用其他模型近似代替. 常采用以下几种方法:

1) 去掉一些变量.

2) 合并和细分一些变量.

3) 改变变量的性质.

4) 改变变量之间的函数关系.

5) 改变约束条件.

形象模型(iconic model) 见“模型”.

抽象模型(abstract model) 见“模型”.

确定性模型(deterministic model) 见“模型”.

随机性模型(random model) 见“模型”.

模拟(simulation) 亦称仿真. 模型论术语. 是借助于计算机、模型或其他装置对实际系统的状况或过程所作的抽象和近似的描述. 通过模拟可以对实际系统做出直接的了解和分析. 特别是一些难以进行实际试验的大规模社会经济活动和军事活动等,都可以通过模拟的方法获得接近实际的重要信息.

模拟模型(simulation model) 一组对某事物的规则和关系的描述. 这种描述可以用思维规则来完成,称之为思维模拟模型;可以用语言的陈述来描述,称之为陈述模拟模型;也可以用方程解析式或图来描述,称之为数学模拟模型. 模拟模型可以从不同的角度分类. 从系统的状态特性看,可分为确定性模型和随机性模型;从系统的时间特性看,可分为离散性模型和连续性模型;从模型的结构形式看,可分为简单式模型、串联式模型、扩展式模型和分析式模型;从运筹学的应用范围看,可分为存贮模型、排队模型、预测模型、更新模型和训练模型等.

仿真(simulate) 即“模拟”.

计算机模拟(computer simulation) 亦称计算机仿真. 一种对实际问题求数值模拟解的技术. 它将一个复杂的客观系统的结构和行为模型化,并通过计算机进行动态的模仿或表演,以安全和经济的方式获得系统或过程的数值结果;然后,经过对比、分析和研究,预测、认识和评价该系统的行为效果,从而为决策者提供可靠的依据. 计算机模拟是 20 世纪 50 年代以来,随着计算机科学和系统科学的发展而产生的一种新的科学方法. 它广泛地应用于经济、军事和科学实验的决策分析中.

计算机仿真(computer simulation) 即“计算机模拟”.

军事运筹学

军事运筹学(military operations research) 运筹学的一个分支. 是运用数学方法和现代计算技术,研究军事活动中的数量关系,从而为军事领域的正确决策提供数量依据的一门学科. 它是现代军事科学的一个新的重要组成部分. 运筹学的形成和发展与军事活动是分不开的;运筹学的英文“operations research”的直译就是作战研究. 从这个意义上说,军事运筹学和运筹学是一个意思. 在第二次世界大战以后,由于经济建设的需要,特别是电子计算机的出现,促使运筹学这门新兴学科迅速发展,理论上日臻完善,应用更加广泛. 它不仅用来研究和解决军事作战问题,而且渗透到社会科学、经济建设、交通运输、能源开发、企业管理等各个领域. 运筹学已经成为现代科学中的一个重要组成部分,而军事运筹学则发展成为运筹学中的一个独特领域. 军事运筹学在第二次世界大战中,主要被用来研究当时武器装备的有效使用和对局部战术问题进行研究. 现代军事运筹学不仅对现有武器装备和局部战术问题提供最优决策方案,而且还借助于电子计算机、自动控制、模拟等先进技术和方法,对未来战争中的武器装备、作战方式和战略策略问题进行预测、计划和评估.

兰彻斯特方程(Lanchester equation) 一种描述交战双方消耗量的微分方程. 它是在第一次世界大战期间,由英国的汽车工程师兰彻斯特(Lanchester, F. W.)提出的. 他通过对古代战争和现代战争中双方军事力量交战损失的研究与分析,提出了线性定律和平方定律,并分别用微分方程描述交战双方的战斗力消耗情况. 兰彻斯特方程常被用来预测战斗的结果和交战双方的损失情况.

线性定律(linear law) 亦称兰彻斯特线性定律. 反映交战双方消耗率的一种规律. 分为第一线性

定律和第二线性定律. 设 $B=B(t)$ 表示在任意时刻 $t(\geqslant 0)$ 蓝方的兵力, $B_0=B(0)$ 是蓝方初始时刻的兵力, $R=R(t)$ 表示在任意时刻 t 红方的兵力, $R_0=R(0)$ 是红方初始时刻的兵力, β 表示蓝方单位时间内兵力损失的数量, ρ 表示红方单位时间内兵力损失的数量.

1. 第一线性定律. 当交战双方使用相似的武器、单兵作战、直接瞄准射击(如古代冷兵器作战、在狭窄的空间里一一格斗等), 双方平均消耗率为一常数, 则兰彻斯特第一线性定律为

$$\begin{cases}\dfrac{dB}{dt}=-\beta,\\ \dfrac{dR}{dt}=-\rho.\end{cases}$$

2. 第二线性定律. 当交战双方进行远距离间接瞄准射击(如炮战)、火力集中在对方战斗单位的集结地区、一方的损失率与另一方射击单位的数量成正比, 且与己方战斗单位的密度有关时, 则可使用兰彻斯特第二线性定律:

$$\begin{cases}\dfrac{dB}{dt}=-\beta RB,\\ \dfrac{dR}{dt}=-\rho BR.\end{cases}$$

在任意时刻 t, 蓝方与红方的力量之比为

$$\frac{B}{R}=\frac{B_0}{R_0}\exp[-\rho B_0(K-1)t],$$

式中 $K=\beta R_0/\rho B_0$. 当 $K<1$ 时, 蓝方胜; 当 $K>1$ 时, 红方胜. 状态方程为: $\rho(B_0-B)=\beta(R_0-R)$.

平方定律(square law) 亦称兰彻斯特平方定律. 反映交战双方消耗率的一种规律. 在线性定律中, 总假定是一对一地进行战斗, 双方的消耗率是常数. 但在现代战争中, 常常采用集中兵力、兵器, 攻击敌方薄弱环节的战略战术, 使一方造成在局部上的数量优势, 而使在数量上处于劣势的另一方无力进行还击. 在这种情况下, 双方的消耗率就取决于某一时刻双方参加战斗的兵力、兵器数的平方, 即兰彻斯特平方定律:

$$\begin{cases}\dfrac{dB}{dt}=-\beta R,\\ \dfrac{dR}{dt}=-\rho B,\end{cases}$$

式中 $\beta,\rho,t>0$, $B\leqslant B_0$, $R\leqslant R_0$. 解得

$$\rho(B_0^2-B^2)=\beta(R_0^2-R^2).$$

当 $\rho B_0^2>\beta R_0^2$ 时, 蓝方占优势; 当 $\rho B_0^2<\beta R_0^2$ 时, 红方占优势.

对数定律(law of logarithm) 兰彻斯特方程的推广. 在大规模战斗中, 当交战双方各自投入的兵力(兵器)超过 15000 人(件)时, 应用线性定律和平方定律就不合适了, 这时需要使用下面的方程:

$$\begin{cases}\dfrac{dB}{dt}=-\beta B\ln R,\\ \dfrac{dR}{dt}=-\rho R\ln B.\end{cases}$$

在一定条件下它可简化为

$$\begin{cases}\dfrac{dB}{dt}=-\beta B,\\ \dfrac{dR}{dt}=-\rho R,\end{cases}$$

从而有

$$\begin{cases}B=B_0\exp(-\beta t),\\ R=R_0\exp(-\rho t).\end{cases}$$

这表明, 一方战斗兵力的损失状况, 取决于他们自己最初的兵力及消耗率常数 β 和 ρ, 而对于对方兵力的依赖程度是很小的.

蒙特卡罗法(Monte Carlo method) 亦称统计试验法. 一种随机模拟方法. 它是以摩纳哥著名的赌城蒙特卡罗(Monte Carlo)命名的借助于概率论的数学模型而解决问题的一种方法. 这种方法既可以解决一些概率型的问题, 又可以解决一些非概率型的问题. 统计试验法以电子计算机为工具, 已得到了广泛的应用和发展.

射击效率评定(efficiency evaluation of fire) 军事运筹学中的一种效率评估方法. 是武器射击理论的一个组成部分. 它应用概率论等数学方法, 确定武器技术性能对射击效率的影响, 比较各种射击方式和战术方法的效率, 从而对武器的射击效率做出正确的评价, 拟定出最有效的射击方式和战术方法. 它研究的内容包括: 对单个目标射击的效率, 对目标群射击的效率, 对面目标射击的效率及在对抗条件下射击的效率等.

射击效率(efficiency of fire) 军事运筹学术语. 是用数量来刻画某种武器射击结果与预定射击任务的符合程度的一个指标.

撰 稿 王长钰 王金德 刘 克 刘家壮 杨家荣 李荣生 宋增民 张亚光 原文志 胡毓达 董相成 谢继国

审 阅 马仲蕃 刘德铭 杨德庄 林元烈 徐光辉 越民义 韩继业 程 侃 谢力同 颜基义

系　统　理　论

系统理论(system theory)　阐述系统的属性、运动、演化规律的概念体系.系统泛指任何由若干部件或要素按一定结构组成,并具有一定功能的集合体.其各部件可以具有不同质的内容和运动规律.系统理论研究的则是由部件之间相互作用的结构性质如何确定其整体功能的普遍规律性问题.已有多种数学工具可用来对这种规律进行一般的、严格的描述,并得到许多理论的和实用的结果.但总的说来,越是抽象、概括的描述,其具体内涵就越少;而越特定、越详尽的理论,其实用性越强,外延也就越窄.由此形成了不同抽象层次上的系统理论,即从最一般的基于语言或集合描述的公理化系统理论,到具一定代数结构或函数空间结构的抽象系统理论,直到用常微分(或差分)方程、代数、图论或离散事件过程描述的更加面向实际问题的系统理论以及相应的一整套用于工程实际的行之有效的技术和方法.

描述事物相互作用随时间演化过程的动态系统理论在数学和物理中早已有之.但首次完整地提出系统概念并明确地指出其在物理、化学、生物及人类社会等领域中的具体应用的,当推 20 世纪 20 年代,奥地利学者冯·贝塔朗菲(von Bertalanffy,L.),在其后发表的"关于一般系统论"的一系列论文以及 1968 年出版的更完整的专著《一般系统论》中,他针对科学研究日益专门化、只注重细节分析的还原主义趋势,提出了从整体、联系上来把握系统全局功能的科学思想.他指出,一般系统论的目标在于探索源于不同学科的形形色色系统所共有的一般规律,而这种规律是客观存在的.由于数学是描述这些跨学科规律的恰当工具,他认为"系统论主要是数学的一个分支".对系统普遍规律的具体描述涉及许多数学分支,有的甚至还难以给出精确和定量的描述,但"不能用数学表述的地方,系统思想仍有价值".书中还提出了系统普遍具有的层次结构形式、有序性和预决性、在流动中维持其准稳态结构的开放系统等基本特征,及其在各类具体系统中的表现等.这些均被公认为系统论的奠基性工作.

与贝塔朗菲几乎同时提出的控制论、信息论和博弈论也都具有跨学科的性质,并且从不同的侧面深化了对一般系统的认识,对系统理论的发展和应用起了非常重要的推动作用.1967 年,比利时学者普利高津(Prigogine,I.)提出关于远离平衡态的开放系统在与外界交换能量、物质、信息的条件下形成并维持有序的耗散结构的理论.德国学者艾根(Eigen,M.)于 1971 年提出的超循环理论,探讨了生命起源与进化中的自组织过程.德国学者哈肯(Haken,H.)于 1976 年建立的协同学,研究由大量子系统强相互作用、"协同动作"而形成复杂系统的宏观有序结构的机制和规律.由美国气象学家洛伦兹(Lorenz,E.N.)于 20 世纪 60 年代研究大气运动时首次发现.其后由美国李天岩和约克(Yorke,J.A.)于 1975 年及费根包姆(Feigenbaum,M.J.)等人于 1979 年建立的系统混沌运动的理论,揭示了由微观的确定性规律产生貌似随机的混沌运动的过程以及由微观混沌运动形成宏观新形式有序结构的机制.大量研究表明,这些理论在不同程度上说明了物理、化学乃至生物和人类活动系统中的复杂运动规律及系统结构与功能的关系,丰富和深化了一般系统论的内容.

与此相应,在数学上提出的分岔和突变理论、微分动力系统和分形分维理论等,为上述系统规律研究提供了有力的工具.美国米沙诺维奇(Mesarovic,M.D.)等人基于集合论的一般系统描述框架、卡尔曼(Kalman,R.)等人的代数系统理论以及由某些控制问题推动的函数空间中的系统理论,均以严格的数学形式阐明了系统通过控制与观测,而与外界相互作用时的结构性质.这些工作已经成为现代控制理论的基础,并使之发展成为促进现代科学技术发展和社会进步的重要的工程技术领域.与此同时,系统工程面向更多样化的各类系统的分析、设计和运行管理问题,也得到了蓬勃的发展.它的应用已涉及社会、经济、军事、人口、生态、资源以及工农业和科技等方面的组织、管理和优化,可以说已经深入到与人们日常生活有关的几乎所有领域.系统思想、系统理论及其有关的工程技术已经产生了广泛的影响,得到了全社会的关注.

随着系统科学的深入发展和广泛应用,人们不断面临着新的挑战.从系统理论发展的历史可以看出,所研究的系统日益庞大、日趋复杂一直是主要的驱动因素.从弹簧振子、简单电路,到稍复杂一些的运动物体、机电控制,人们可以建立足够精确的微分方程模型,并用已有的数学工具进行彻底的分析,现在它们被称为简单系统.随机干扰和多输入-多输出的交叉耦合问题可能是当时遇到的主要挑战,20 世纪 60 年代发展起来的以状态空间线性多变量系统滤波和控制为基石的现代控制理论较好地解决了这些问题.随后遇到的困难,一方面是在工业、生物等

领域难以建立精确模型,另一方面则是用于电力、交通等大系统时,其信息采集、传输和处理代价过大和可靠性的问题.前者推动了模糊控制、智能控制等分支的发展,后者则导致20世纪70年代大系统理论研究的一股热潮.进入20世纪80年代,人们进一步认识到,上述系统,特别是加上人类决策过程和计算机广泛应用,所呈现的复杂现象不仅在许多方面与生物学家、物理学家所研究并揭示的“复杂性”极其相似,而且在大多数情形下更有过之,甚至无法用现有的单一的精确模型来描述和解决.变量和元素个数的巨量增加导致了巨系统的概念.在耗散结构、协同学等自组织理论中,物理学家用以展示宏、微观相互作用机制所用的一对(两层次)宏、微观关系的模型(钱学森称之为简单巨系统),在现实生活中可能要拓广到更多的层次.子系统种类多、关系复杂,意味着需要用多种不同质的模型来描述,甚至对同一系统或子系统的不同侧面也要用不同的模型,然后还要将它们集成起来.随着信息技术进步,人们在实践过程中不断把“复杂”的东西(一旦弄清)转化为“简单”的,把难以处理的“巨量”的东西转化为可处理的,并加以实现.在系统理论的发展中,充满了大与小、简单与复杂、宏观与微观、定性与定量等概念的对立与转化.由中国科学家钱学森提出的对开放的复杂巨系统的从定性到定量综合集成法指明了系统理论的发展方向.人们正是沿着这一方向,在不断实践与认识的长河中,推动系统理论的发展,使之在科技进步和社会发展中发挥越来越大的作用.

一般系统理论

一般系统论(general system theory) 关于一般系统的概念、性质和研究途径的理论.由奥地利学者冯·贝塔朗菲(von Bertalanffy,L.)于1937年首先提出.在题为“一般系统论”的奠基性著作中,他建立了探索源于不同学科领域的一般系统的共同概念、性质和规律的横断交叉学科分支.他指出原有生物学研究中的还原论思想的缺陷,提出了总体论和还原论结合的系统论思想.在其奠基性专著和一系列论文中,他全面论述了一般系统概念、封闭和开放系统、信息和熵、系统组织、动力学和演化、反馈和自稳定等系统的普遍性质,以及它们在物理、化学、生物、医学、心理学和社会学等方面的应用.

寻的系统(goal-seeking system) 一类由其内在目标导向而不单是由刺激/响应规则来确定其行为的系统.它所寻求的目标可能是优化、决策、控制、问题求解、目标跟踪、模式识别等,从而可表示更广泛的一类系统,特别是与人类活动有关的社会系统.建立寻的系统必须给定系统目标,说明期望决策的含义,并描述决策的过程.形式上它可表示为7元组 $GS=(M,U,V,Y,P,G,\phi)$.其中,M 为备选决策集合;U 为决策者内在不确定性集合;V 为价值评估集合,通常具有可比较优劣的半序结构;Y 为可能输出的集合;关系 $P\subset M\times U\times Y$ 表示系统内部受控过程的行为;$G:M\times U\times Y\to V$ 是用来评估系统行为的品质函数;ϕ 为出现冲突或不确定性时的决策原则.

一般系统的数学定义(mathematical definition of general systems) 关于一般系统的一种形式化描述.从“系统是由其若干组成部分相互联系、相互作用而形成的统一体”这样最一般的概念出发,按其抽象的层次不同,可以有多种不同可操作性的数学形式的定义.为了说明一个系统,最起码可以用自然语言来说明系统各要素及其相互作用和联系.例如,各语句中的名词成分往往代表系统的各种对象或要素,而把它们连结起来的谓语或附加成分则用来说明其间关系.这可以说是最一般的、用自然语言对系统进行的定义.

随后是用形式语言进行定义.这时,一个系统是某形式语言中规定好的若干合法句子所构成的一个子语言.用集合论可以把上述概念进一步形式化如下:系统中各种对象或要素可按其性质及相互作用时的地位而归并成若干集合 $V_1,V_2,\cdots,V_n$,它们之间的相互作用可定义为这些集合之间的关系.按照集合论,它们无非就是直积集合 $V_1\times V_2\times\cdots\times V_n$ 的某个子集合.由它们又可通过关系间的种种运算而形成更复杂的系统结构,并由此导出系统的许多基本性质,诸如状态演化动力学、输入-输出响应、实现问题、能控能观性、稳定性以及系统的分解、互连、解耦等(参见“基于集合论的一般系统数学定义”).

为了得到进一步的、具有实用价值的结果,自然要涉及各种系统独特的结构性质.这反映在系统定义中,对象集合及其运算可以具有不同的结构,从而形成相应的不同描述和理论.例如,对象元素在所定义的运算下可能形成某种代数结构,相应地可用抽象代数来描述系统.如有限自动机的代数理论,或卡尔曼(Kalman,R.)等人建立的“多项式环上的模”上的线性系统理论在解决实现问题方面取得的进展;又如用极大代数可以描述排队、服务等离散事件过程中事件发生时刻之间的关系(参见“极大代数”).

对象集合 V_i 所具有的拓扑结构决定了所采用的数学工具及理论的面貌.有人用一般拓扑、同调拓扑或代数拓扑等工具来研究一般拓扑空间上的系统.当 V_i 是某些函数空间时,可以用泛函分析的方法来研究,可以把分布参数系统、随机系统等无穷维情形和普通有穷维系统的观念统一起来,得到许多实质性的结果.进而当各 V_i 为通常的向量空间并引入时间以描述其动态过程时,可以用微分方程、差分

方程等成熟的数学工具来研究.无论对线性还是非线性系统,都有许多定性和定量的方法,可以得到丰富的结果.这是一般系统论中结果最丰富、应用最广泛的部分.

对于开放系统或具有输入-输出作用的系统,它的各要素可以分为输入 $X=(V_1,V_2,\cdots,V_m)$ 和输出 $Y=(V_{m+1},V_{m+2},\cdots,V_n)$ 两部分.它们之间的关系还要引入一组状态变量 Q 来描述(参见"基于集合论的一般系统数学定义").

在自然科学和工程技术中,常见的动态系统可以用常微分方程组描述.设输入 X 的分量为 $x_1,x_2,\cdots,x_m$,输出 Y 的分量为 $y_1,y_2,\cdots,y_p$,而状态 Q 的分量为 $q_1,q_2,\cdots,q_r$,则 X 和 Q 的关系为

$$\begin{cases}\dfrac{\mathrm{d}q_1}{\mathrm{d}t}=f_1(q_1,q_2,\cdots,q_r;x_1,x_2,\cdots,x_m),\\ \dfrac{\mathrm{d}q_2}{\mathrm{d}t}=f_2(q_1,q_2,\cdots,q_r;x_1,x_2,\cdots,x_m),\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \dfrac{\mathrm{d}q_r}{\mathrm{d}t}=f_r(q_1,q_2,\cdots,q_r;x_1,x_2,\cdots,x_m).\end{cases}$$

而输出 Y 可表示为一般的函数关系:

$$\begin{cases}y_1=h_1(q_1,q_2,\cdots,q_r;x_1,x_2,\cdots,x_m),\\ y_2=h_2(q_1,q_2,\cdots,q_r;x_1,x_2,\cdots,x_m),\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ y_p=h_p(q_1,q_2,\cdots,q_r;x_1,x_2,\cdots,x_m).\end{cases}$$

当 f_i, h_j 均为线性函数时,上述方程就是线性输入-输出动态系统的模型.对此已有相当完善的理论.更复杂的动态系统模型可能要用偏微分方程、带时滞的微分方程、积-微分方程描述.另一方面,当 X,Y,Q 诸元素均为离散的符号量时,又可能用有限状态机等模型来描述(参见"离散事件动态系统").

基于集合论的一般系统数学定义(mathematical definition of general system based on set theory) 关于一般系统的一种形式化描述.用最一般的集合论概念可以把一般系统定义为:在非空集合 $\times\{V_i: i\in I\}$ 上定义的一个关系 S:

$$S\subset\times\{V_i: i\in I\}, \tag{1}$$

其中"×"表示{·}中各集合的直积,I 为指标集,当 i 为有限集时,(1)式可改写为

$$S\subset V_1\times V_2\times\cdots\times V_n.$$

定义中各 V_i 实际上就是系统的各组分对象集,系统描述应当说明它们之间的关系.特别指出,若 $x\in V_1$, $y\in V_2$,则函数关系 $y=f(x)$ 即为子集:

$$S=\{x,y=f(x)\}\subset V_1\times V_2.$$

一般的关系可以表示多值的函数,其含义比函数更广.输入-输出系统定义为:当系统各组分元素可划分为输入 $X=\times\{V_i: i\in I_x\}$ 和输出 $Y=\times\{V_j: j\in I_y\}$ 两部分,即

$$I_x\cup I_y=I,\ I_x\cap I_y=\varnothing,$$

则系统 $S\subset X\times Y$ 称为输入-输出系统.直观上,输入常代表外界对系统的刺激和影响,输出表示系统的反应及其对外界的影响.一般地,输出并非由输入惟一确定(例如,还依动态系统的初值而变),S 只是一个关系.如果输出可由输入惟一确定,即 S 是个函数,则称之为函数型系统.

系统的状态指的是为完全确定系统输出所需的一组信息,它可定义为:若 Q 为某集合,而 R 为函数,$R:(Q\times X)\to Y$,且有

$$(x,y)\in S\Longleftrightarrow(\exists q\in Q)[R(q,x)=y], \tag{2}$$

则 Q 为系统 S 的状态集,$q\in Q$ 为系统的状态.在实际应用中,往往还要求 Q 是具有上述性质的最小集合.当 X,Y 均为某域 $\mathscr{A}$ 上的线性代数,而 S 是 $X\times Y$ 中的非空线性集,亦即对 $s,s'\in S,a\in\mathscr{A}$,均有 $s+s'\in S,as\in S$,则 S 是个线性系统.可以证明,S 是线性系统的充分必要条件是可以找到状态集 Q 亦为线性代数,而且函数 R 可分解为

$$y=R(q,x)=R_1(q)+R_2(x), \tag{3}$$

其中 R_1,R_2 均为线性映射,分别称为系统的状态响应和输入响应.

在刺激与响应、因与果之间总有时间滞后,系统各部分间或其与环境的相互作用是一种动态过程,只有在时间演化过程中才能充分显示系统运动的各种复杂形态.为此,首先要引进时间概念:时间集合 T 可以是任何线性有序集,其中的序关系常用"≤"来表示.系统的输入、输出及状态均随时间而变,从而 $X\subset A^T$, $Y\subset B^T$,$Q\subset C^T$,这里 A^T 表示定义于时段 T 上而取值于集 A 的时间函数,B^T,C^T 亦然.在这样的 X 和 Y 上定义的系统,称为时间系统.但时间系统要成为动态系统,还要满足一些附加条件:

1. 存在初始状态集 Q_0,使得输入-输出关系 $(x,y)\in S$ 可以用映射 $\rho_0: Q_0\times X\to Y$ 来表示,亦即

$$(x,y)\in S\Longleftrightarrow(\exists q\in Q_0)[\rho_0(q,x)=y].$$

2. 任何时刻的状态集 Q_t 可以定义为 S 在 T 的子集 $T_t=\{\tau: \tau\geqslant t,\ \tau\in T\}$ 上的"限制系统"S_t 的初始状态集,亦即

$$(x_t,y_t)\in S_t\Longleftrightarrow(\exists q\in Q_t)[\rho_t(q,x_t)=y_t],$$

其中,x_t 是时间函数 x 在 $\tau\geqslant t$ 时段的限制,y_t 亦然.而 ρ_t 称为 t 时刻的系统响应函数.所有 ρ_t 的族称为响应族,记为 $\bar{\rho}=\{\rho_t=Q_t\times X_t\to Y_t,\ \forall t\in T\}$.

3. 在不同时刻 t 和 t'(设 $t<t'$)的状态 q_t 和 $q_{t'}$ 间存在状态演化关系 $\varphi_{tt'}: q_t\times X_{tt'}\to q_{t'}$,使得

$$q_{t'}=\varphi_{tt'}(c_t,x_{tt'}),$$

其中 $x_{tt'}$ 表示时间函数 $x(\tau)$ 在 $t\leqslant\tau\leqslant t'$ 时段的限制.所有 $\varphi_{tt'}$ 的集合称为状态演化映射族,记为

$$\bar{\varphi}=\{\varphi_{tt'}: q_t\times X_{tt'}\to Q_{t'},\ \forall t,t'\in T,\ t<t'\},$$

而 φ 还必须满足动态系统的相容性条件或半群条件：

1) $\rho_t(q_t, x_t)|T_{t'} = \rho_{t'}(\varphi_{tt'}(q_t, x_{tt'}), x_{t'})$，等式左边即 $y_t|T_{t'}$，表示时段 $\tau \geqslant t$ 上的函数 y_t 在 $T_{t'}$ 上的限制.

2) $\varphi_{tt'}(q_t, x_{tt'}) = \varphi_{t''t'}[\varphi_{tt''}(q_t, x_{tt''}), x_{t''t'}]$，$t', t'', t \in T$ 且满足 $t < t'' < t'$.

条件 1)表示初态 q_t 和输入 x_t 所确定的输出 y_t 在时段 $T_{t'}$ 上等于由初态 $q_{t'}$ 和输入 $x_{t'}$ 所确定的输出 $y_{t'}$. 条件 2)表示由 q_t 出发在 $x_{tt'}$ 输入下达到的 $q_{t'}$（方程左端）等于先用 $x_{tt'}$ 输入达到 $g_{t''}$，再在 $x_{t''t'}$ 作用下于 t' 时刻达到的状态. 这种半群性质或结合性质反映出在任一时刻 t''，状态 $q_{t''}$“总结”了此前所有历史信息，而其后的演化将只与此后 $q_{t''}$ 和 $x_{t''t'}$ 有关. 这个性质对于理解“状态”的本质以及对动态系统的分析与控制都有很重要的意义. 有了上述概念，就不难定义一般动态系统的实现问题、具控制输入时的能控性问题、系统的各种稳定性概念，诸如李亚谱诺夫(对初值的)稳定性、输入-输出稳定性、结构稳定(鲁棒)性等. 还可以在如此一般的系统定义下研究子系统的串联、并联和反馈连结等.

通有性(generic property)　一种系统属性. 指对几乎所有参数均满足的系统性质；亦可描述为：在参数空间中，对应不具该性质的参数总体测度为零.

决策系统(decision system)　用于决策目的的系统. 指由其决策目标来定义其行为导向的系统(参见“寻的系统”). 决策系统通常应包括根据形势可调整的确定决策目标的机制、数据信息采集和处理机制，包含已有模型、规则和算法的知识库，以及便于人类交互和操作的友好人机界面.

控制系统(control system)　具有控制机制的系统. 指通过控制作用使得对象行为在不确定条件下符合并保持预定功能的系统. 通常的预定功能指系统状态演化符合预定的轨道要求，或使某种品质指标达到最佳. 所谓不确定条件主要为环境变化或噪声干扰，以及系统本身部件特性的变化或失灵. 采用的控制方案包括按行为要求输入指令的开环控制，以及采集关于干扰和误差的信息经适当处理后反馈输入的闭环控制.

叠加原理(superposition principle)　区分线性与非线性的一条基本准则. 令 x 为系统的输入变量，y 为系统的输出变量，$f()$ 为输出对输入的响应函数 $y = f(x)$. 该系统满足叠加原理，指以下两个条件同时成立：

1. 可加性. $f(x_1 + x_2) = f(x_1) + f(x_2)$.

2. 齐次性. $f(kx) = kf(x)$ (k 为任何常实数).

凡同时满足可加性和齐次性要求的是线性系统，至少一个要求不满足的是非线性系统.

线性系统(linear system)　满足叠加原理的一类系统. 指输入、状态、输出之间的静态或动态关系均符合叠加原理的系统. 任何因果关系满足叠加原理指的是：两个原因相加所导致的结果即为它们分别导致结果之和；同时原因的倍数产生的结果亦为原结果的相同倍数. 常见的线性系统可用线性常微分方程、偏微分方程描述或差分方程描述. 由于已有较完善的定性和定量分析方法，对线性系统理解得较为透彻，并往往成为更广的各类系统分析方法的基础(参见本卷《控制理论》同名条).

非线性系统(nonlinear system)　不满足叠加原理的一类系统. 指其特性不能用线性系统模型描述和研究的系统. 尽管逻辑上任何不满足叠加原理的系统都是非线性系统，但实际上只有那些系统行为特征无法用线性模型逼近的系统，才算是本质的非线性系统.

双线性系统(bilinear system)　一类特殊的非线性系统. 其状态空间描述的微分方程右端为状态 x 和输入 u 的双线性型，即

$$\dot{x} = Ax + Bu + uFx,$$

其中 F 为适当维数的矩阵. 这类系统适于描述化学反应、生物繁殖、量子力学过程、社会经济和人口等的动态演化过程，有重要的应用背景. 由于其“增益”随输入 u 而变，具有一定的自适应性，亦被称为变结构系统. 双线性系统的研究始于 20 世纪 70 年代，其主要的理论结果包括：对 u 的幅度受限情形，已得到判别稳定性的频域准则；得到系统完全能控的充分必要条件，表明较一般非线性系统有更强的能控制性；用李亚普诺夫稳定性理论求出其状态反馈镇定解；以及用极大值原理和动态规划求解其最优控制问题的若干结果.

变结构系统(variable structure system)　系统结构可改变的系统. 通常为使系统具有更强的功能而设计的具灵活可变结构的系统. 其结构可按系统状态或其他信号加以改变. 常见的如变结构控制系统. 此外，在早期文献中亦将双线性系统归类为变结构系统.

代数系统理论(algebraic system theory)　数学系统理论的一个分支. 它通过对大量原子系统，赋以函数变换和各种互连结构以形成复杂多样的系统特性，并进而研究系统变量集合及其运算的代数性质，以及由此形成的系统的代数结构. 它可突出不同类型系统的共同特点，从而扩大系统理论的应用范围. 其主要内容包括定义在一般的集合和映射运算上的半群、环、模上的系统的各种模型及其主要性质，在理论上和应用中都有重要意义.

为给定函数变换环节，必须说明其定义域、值域和变换关系，而互连结构则包含通常的串联、并联及

其形成的反馈结构等.通过大量原子系统的组合、叠加,可以得到极为复杂的具多层次结构的大系统.特别是采用群、环、域、模等代数系统和相应的同态变换,可以得到多种不同的系统理论以描述各种事理过程.例如,集同态系统、半群同态系统、模同态系统等,从而可大为拓广系统理论的应用范围.

半群同态系统(semi-group homomorphic system)　一类基于半群结构定义的系统.指系统各变量集合具有半群结构而输入/状态迁移函数 f 和输出观测函数 g 均为半群同态映射的集动态系统(参见“集动态系统”).若$(S,\diamond,e_S)$和$(T,\Box,e_T)$均为具单位元的半群,e_S,e_T分别为它们的单位元.设 f 为由 S 到 T 的映射,若对 S 中任意元素 s_1,s_2,均有

$$f(s_1)\diamond f(s_2)=f(s_1)\Box f(s_2),$$

则称 f 为半群同态映射.在半群同态系统中,由于 f 和 g 分别为 $X\times U\to X$ 和 $X\times U\to Y$ 的映射,首先需要定义 $X\times U$ 的半群结构.可以证明,在适当定义下,一个半群同态系统可用下列动态方程来描述:

$$x(k+1)=Ax(k)\diamond Bu(k),$$
$$y(k)=Cx(k)\Box Du(k).$$

最常见的线性定常系统之所以具有这种形式,也正由于它是半群同态系统.由于这种比较简单的标准形式,使得半群同态系统既比常见的域上系统更为广泛,又可得到十分丰富的结果,从而在社会、心理、离散事件等非域上动态系统研究中有重要的应用.

模同态系统(module homomorphic system)欧氏空间系统的一种推广.群同态系统引进 R 环结构而成的 R 模上的动态系统.对给定的环

$$R=(R,*,e,\Delta,1_R),$$

设半群同态系统(U,X,Y,f,g)中 U,X,Y 均为交换群,并在其上均定义了 R 环结构,即映射 $h:R\times G\to G$(G 泛指 U,X,Y),从而构成了 R 模.此外,还要求映射 h 满足以下条件:

$$f(r,g_1)\Box f(r,g_2);$$
$$f(r_1*r_2,g)=f(r_1,g)\Box f(r_2,g)f(r_1\Delta r_2,g)=f(r_1,f(r_2,g));$$
$$f(1_R,g)=g.$$

这些条件实际上是下列欧氏空间中向量与实数乘法之间的关系的一种推广:

$$\alpha\cdot(x+y)=\alpha\cdot x+\alpha\cdot y;$$
$$(\alpha+\beta)x=\alpha x+\beta x;\ (\alpha\cdot\beta)x=\alpha(\beta x);$$
$$1\cdot x=x.$$

因此,模的概念相当于环上的向量空间.对于一个模同态系统,若输入为$(u_0,u_1,\cdots,u_i)$,状态初值为 e_x 时,系统状态为

$$\left(\mathop{\Box}_{j=0}^{i}a^{i-j}b\right)(u_0,u_1,\cdots,u_i)=\mathop{\Box}_{j=0}^{i}a^{i-j}bu_j$$

可以证明映射

$$\mathop{\Box}_{j=0}^{i}a^{i-j}b$$

是一个模同态.模同态系统的重要应用之一是动态系统的实现理论.

坡模上的同态系统(homomorphic system on the incline module)　在交换半群上引入半格结构而形成的坡模上的动态系统.所谓坡(slop 或 incline)是指非空集合 P 上定义了两种二元运算“+”和“·”,使对 P 中任意元素 a,b,c,均满足:

1.$(P,+)$是一个半格.

2.$(P,\cdot)$是一个交换半群.

3.分配律:$a\cdot(b+c)=a\cdot b+a\cdot c$.

4.吸收律:$a\cdot b+a=a$.

进而,交换半群$(M,+)$称为坡 P 上的模(简称坡模),是指对 $x,y\in M$ 及 $a,b\in P$,存在映射

$$P\times M\to M:(a,x)\to ax,$$

使其满足:

1) $a(x+y)=ax+ay$.

2) $a(bx)=(ab)x$.

3) $(a+b)x=ax+bx$.

4) 若 P 具有单位元 1,则 $1x=x$.

5) 若 P 具有单位元 0,则 $0x=\theta$,其中 θ 为 M 中的加法单位元.

类似地,有坡模上的同态系统的一般形式:

$$x(k+1)=Ax(k)+Bu(k),$$
$$y(k)P=Cx(k)+Du(k),$$

并可得到这类系统的一系列性质.由于坡集合的特性,这类系统适于描述特殊的离散状态或心理现象等,有一定的应用价值.

前瞻系统(anticipatory system)　一类具有预见能力的系统.指当前行为或输入以某种形式取决于其未来状态的系统.这表面上违反物理因果律的现象,实际上只是系统可以利用已有的物理规律和观测数据来预见其将来行为.对于系统某些部件具有时间滞后的情形,利用有前瞻或预测部件,可以改进控制品质.此外,在微分方程和非微分约束相耦合的奇异系统中,可能出现脉冲模,即系统可在瞬间即可做出有限的反应,这实际上也是一种前瞻行为.

函数型系统(function-type system)　一类可用输入-输出间的函数关系来定义的系统.一般系统定义中 S 不仅是关系,而且是函数 $S:X\to Y$ 的情形,这时输出 Y 由输入 X 完全决定(参见“基于集合论的一般系统数学定义”).

抽象线性系统(abstract linear system)　线性系统概念的一种推广.若 $\mathscr{A}$ 为任一域,X,Y 为 $\mathscr{A}$ 上的线性代数,关系 $S\subset X\times Y$ 设为非空,此外设:

1. $s,s'\in S\to s+s'\in S$;

2. $s\in S,\alpha\in\mathscr{A}\to\alpha s\in S'$;

则 S 为一线性系统的充分必要条件为：存在全局响应函数 $R: C\times X\to Y$ 使得：C 是 $\mathscr{A}$ 上之线性代数；存在线性映射 $R_1: C\to Y$；$R_2: X\to Y$，使对所有 (c,x)，有 $R(c,x)=R_1(c)+R_2(x)$.

利用这一概念，可以把已有线性系统的丰富成果推广而用于若干非传统的问题.

系统实现问题(system realization problem) 系统分析的反问题. 指由系统输入-输出之间的外部关系求其完整的状态空间内部描述的问题. 例如对线性系统，其外部描述可以是传递函数形式，或者是充足的输入-输出数据. 其状态空间描述则归结为确定足够的状态变量和完整的微分或差分方程的系数矩阵. 更一般情形是给定输入-输出映射 $\mathscr{F}: U^*\to Y^*$，其中 U^*，Y^* 分别为输入 U 和输出 Y 的时间函数的集合，求状态空间 X、输入-状态迁移函数

$$\bar{\phi}=\{\phi_{tt'}: X_t\times U_{tt'}\to X_{t'}\}$$

和输出响应函数

$$\bar{\rho}=\{\rho_t: X_t\times U_t\to Y_t\},$$

使其产生的输入-输出关系能够实现或最佳逼近预先给定的 $\mathscr{F}$. 对线性系统，对应于同样输入-输出响应的状态空间模型均相互等价，即可经相似变换相互转换. 状态空间维数最低的实现称为最小实现，可以证明，最小实现的充分必要条件是它必为能控(能达)且能观的.

有限维定常线性系统的实现是最简单的情形，已有若干实用的算法. 对布尔变量(取值 0 或 1)系统，输入-输出映射可以用有限自动机等系统模型来实现. 对给定的非线性输入-输出映射可求其非线性变换或非线性动力学的系统实现，难度更大. 更深入的研究有卡尔曼(Kalman,R.)等人的模论方法和范畴论的方法等.

对非线性系统，由于存在各种不同的表现形式，实现问题通常是针对特定表现形式而提的. 例如，求双线性模型的实现，求伏泰拉级数展开形式和实现等(参见“伏泰拉级数展开系统模型”、“双线性系统”).

最小实现(minimal realization) 状态空间维数最低的系统实现. 详见“系统实现问题”、“状态空间”.

抽象系统理论(abstract system theory) 一般系统的一种形式化理论. 指脱离系统的具体领域内容，通过将其口语化的定义转化为由抽象而精确的数学语言而对一般系统概念和性质进行研究的理论. “抽象”的含义主要指其一般性、广泛性和通用性，从而可使其理论更为简洁和优美. 其具体内容包括：基本系统概念及其形式化描述，系统一般属性，系统主要类型(端口系统、时间系统、寻的系统)，以及系统的复杂性问题.

复杂系统(complex system) 难以用还原论方法处理的各种系统的总括. 泛指规模大、变量多、各部分关系和运动规律复杂或具有较大不确定性的系统. 其准确含义可由对之进行描述和操作的方法加以界定，即它须用处于多个宏观、微观层次上，不同性质的模型来描述、分析、设计和控制. 典型的例子，如包含连续物理变量和逻辑开关控制的混杂系统(参见“混杂动态系统”)，包含人类和机器交互决策的人机系统，定性管理层和可定量化的物质生产过程构成的制造系统、社会经济系统、人类生物和自然环境相互作用的生态系统等.

依待解决问题的不同，对复杂系统需建立不同的模型体系，并用不同的技术手段加以解决. 常用概念和建模体系框架，各种分析方法和技术手段，解题的各种程式和步骤，构成了复杂系统研究的基本方法论.

奈洛德等价(Nerode equivalence) 一种等价关系. 即用以说明用不同状态描述同一输入-输出映射必须满足的条件. 设给定输入-输出映射 $f: X^*\to Y$，其中 X^* 为随时间变化而取值于空间 X 的时间函数集合，令 Ef 为 X^* 上的等价关系：

$$xEfx'\Leftrightarrow f(xz)=f(x'z),$$

对所有 $z\in X^*$. 由于

$$xEfx'\Leftrightarrow fLx=fLx',$$

人们将奈洛德等价类 $[x]_E$ 看成 Qf 的一个元素，Qf 为 f 的最小实现的状态集.

时间系统(time system) 随时间演化的系统. 在一般动态系统中，时间集合 T 可以是任何的线性有序集，其序关系通常用“$\leqslant$”表示. 系统输入、输出均随时间变化，从而输入 $X\subset A^T$、输出 $Y\subset B^T$、$Q\subset C^T$ 分别为取值于集 A,B 和 C 的时间函数，这种系统称为时间系统. 时间系统还须满足一些附加的条件才成为一般动态系统，如可化为初始状态的映射、相容性、半群条件等(参见“基于集合论的一般系统数学定义”).

状态生成函数(state generating function) 一种刻画系统特性的函数. 指由 t 时刻前的输入、输出值确定 t 时刻系统状态的函数法则. 动态系统任何时刻 t 的状态可由初始状态 C_0 和 t 之前的输入 x^t 所确定，也可能由某时刻 $\hat{t}$ 之前的输入和输出所确定. 亦即若 $\bar{\rho}$ 为时间系统 S 的响应族而关系 $\eta^t\subset X^t\times Y^t\times C_t$，则 $(x^t,y^t,c^t)\in\eta^t$ 意味着：

$$(\forall x_t)(\forall y_t)$$

$$[(x^t\cdot x_t,y^t\cdot y_t)\in S\to y_t=\rho_t(c_t,x_t)].$$

当 η^t 为函数，则称 $\eta^t: X^t\to C_t$ 为 t 时刻的状态生成函数.

输出生成函数(output generating function)

一种刻画系统特性的函数. 可由任何给定之输入 x' 和初始状态 C_0 求出任意时刻 $t\in T$ 的输出值 $y(t)$ 的函数关系. 它可形式地定义为 t 时刻状态空间 C', $[t,t']$ 时段的输入值 $\overline{X}_{tt'}$ 和 t' 时刻 $Y(t')$ 值的乘积空间中的一个子集

$$\mu_{tt'}\subset C_t\times\overline{X}_{tt'}\times Y(t'),$$

这种定义反映了多值映射的可能性. 当该关系表示单值映射时, 即为输出生成函数.

因果性(causality) 因果关系的系统论诠释. 在系统论中有两种含义: 其一是由初始响应函数和给定的初态决定系统运动的"非前瞻性"; 其二则为输入决定输出的"过去决定性". 两者通常并用.

非前瞻系统(non-anticipatory system) 一类输出变化不可能先于输入变化的系统. 对时间系统 $S\subset X\times Y$, 若其系统初值响应函数 $\rho_0: C_0\times X\to Y$ 满足以下条件, 则该系统称为非前瞻的:

$$(\forall t)(\forall C_0)(\forall x)$$

$$[x|\overline{T}^t=\hat{x}|\overline{T}^t\to\rho_0(C_0,x)|\overline{T}^t=\rho_0(C_0,\hat{x})|\overline{T}^t].$$

这表示该系统输出的变化不可能先于输入的变化. 注意, 此定义适用于任何时间系统, 而不仅仅是动态系统.

输入-输出系统(input-output system) 用输入-输出关系来描述的系统. 形式化的描述为: 系统 $S\subset V_1\times V_2\times\cdots\times V_n$ 的各组分 V_i 可以明确分为输入集 $X=\times\{V_i: i\in I_x\}$ 和输出集 $Y:\times\{V_i: i\in I_y\}$ 两部分, $X\cap Y=\varnothing$, 于是输入-输出系统定义为 $S\subset X\times Y$, 即为输入-输出的一个关系.

时不变系统(time-invariant system) 亦称平稳系统. 特性不随时间变化的系统. 数学上可精确定义为: 在时间平移变换下保持形式不变的系统. 在用微分或差分方程描述的时不变系统中, 其方程式中不显含时间变量 t.

平稳系统(stationary system) 即"时不变系统".

抽象传递函数(abstract transfer function) 传递函数概念的一种推广. 首先在输入对象 Y 中引入一些结构. 设 V 为某线性代数, $A=V^m$, $B=V^r$, 令 $U\subset V^T$ 为满足下列条件的时间对象:

1. U 是 A 上的线性代数.
2. U 是交换环, 其运算 $*$ 满足: 对 $u,\ u'\in U,\ u*u'=0\Leftrightarrow u=0$ 或 $u'=0$.
3. X 和 Y 可以分别表示为 $X=U^m$, $Y=U^r$.

以上概念的例子如: $T=[0,\infty)$, $V=\mathrm{R}$(实数), $U=C(0,\infty)$ 连续函数族, 对每个 $t\in T$, 定义 $*$ 为卷积运算

$$(u*\hat{u})(t)=\int_0^t u(t-\sigma)\hat{u}(\sigma)\mathrm{d}\sigma,$$

易证卷积运算是交换的, 而且上述条件 1,2,3 均得以满足. $U=C(0,\infty)$ 可以选作为基本时间对象.

设线性输入-响应函数 $y=\rho_{20}(x)$ 定义为

$$y_i=\sum_{j=1}^{m}W_{ij}*x_j\quad(i=1,2,\cdots,r),$$

其中, 权函数 $W_{ij}\in U$, 从而给出用权函数表示动态系统的一种抽象形式. 定义 (u,u') 和 $(\hat{u},\hat{u}')$ 等价之含义为 $u*\hat{u}'=\hat{u}*u'$, 由代数可知 u^2/E 称为商域, 其零元为 $(0,u)$, 幺元为 $[u,u]$. 再定义

$$h(u)=[u*u_0,u_0],$$

其中 $u_0\in U$ 为某固定非零元, 则 h 为满足

$$h(u+u')=h(a)+h(u'),$$

$$h(u*u')=h(u)\times h(u')$$

的一个同态. 将 h 作用于 $\rho_{20}(x)$ 的表达式, 可得

$$h(y_i)=h\Big(\sum_{j=1}^{m}W_{ij}*x_j\Big)=\sum_{j=1}^{m}h(W_{ij})\times h(x_j).$$

这样 ρ_{20} 的抽象传递函数 $TF(\rho_{20})$ 由下列矩阵给出:

$$TF(\rho_{20})=\begin{bmatrix}h(W_{11}) & h(W_{12}) & \cdots & h(W_{1m})\\ h(W_{21}) & h(W_{22}) & \cdots & h(W_{2m})\\ \vdots & \vdots & & \vdots\\ h(W_{r1}) & h(W_{r2}) & \cdots & h(W_{rm})\end{bmatrix},$$

特别在 $r=m=1$ 的情形, 抽象传递函数即为

$$h(W_{11})=\frac{h(y_1)}{h(x_1)}.$$

不难验证, 抽象传递函数概念是通常由拉普拉斯变换导出的传递函数概念的一种推广.

图灵机动态系统(Turing machine as a dynamic system) 一种离散动态系统. 指作为一种动态系统而定义的图灵机. 它可用来描述或实现计算过程或更一般的离散的、逻辑的演化过程, 由英国数学家图灵(Turing, A. M.)于 1936 年提出. 图灵机常被表示为由一个有限自动机控制纸带左右移动, 同时让读写头读出纸带上的字符并加以改写. 这里存贮纸带由无限长的方格串组成, 每方格可写入一个字符, 表示被加工的信息. 控制自动机接收读出的字符、变换自己的状态并发出的改写指令, 这代表了信息处理的计算规则. 除了计算, 它还可对输入的符号串进行识别, 可用于识别形式语言和可计算函数. 图灵还提出了一种更加简洁的通用图灵机, 在概念上完全等同于至今沿用的存入程序式的电子计算机的原理结构. 其后出现的波斯特机、广义图灵机、王浩的 W 机等, 均为图灵机的推广和改进.

计算动态系统(computation dynamic system) 作为动态系统来处理的计算过程. 许多计算过程归结为数据迭代、沿网格传播、搜索与寻优等, 均具有动态系统的各种特点和类似的问题. 利用动态系统的种种性质和丰富成果来研究和改进计算方法已经引起人们的广泛注意.

计算动态过程按预定的规则或算法对初始数据

施加一系列变换，按预定的终止规则结束并输出所得结果. 从而该动态过程是离散时间的、自治的、非前瞻的，可以用系统状态迁移函数来规定，由给定初始值 c 和算法 $\phi_i \in \Phi$，逐步进行演化. 若在有限步数达到平衡态，计算即告完成，这里实际上关心的是有限时间的稳定性问题. 从实用的观点，所谓有限时间也不应该具有过高的复杂度，例如 NP 的计算长度会被认为是不可行的.

一般系统稳定性(stability of general system) 稳定性概念在一般系统情形的推广. 系统行为或后果当出现扰动等原因时得以保持原有要求或连续依赖的性质. 设 $d \in D$ 为引起系统变化的原因，如初始值的误差、外界扰动等；而 $e = F(d) \in E$ 为它对系统影响带来的效果. 这时，最一般的稳定性概念要求，对 d 的小扰动引起 e 的偏差应当不大. 对其严格说明如下：对任意集 V，以 $\sum(V)$ 表示其子集全体，$\Theta_V \subset \sum(V)$ 为某子集类，对任何 $v \in V$，可定义 V 的相对于子集类 Θ_V 的邻域集 $N(v) = \{\alpha : \alpha \in \Theta_V$ 且 $v \in \alpha\}$，这里各 α 均为 v 的邻域，它们属于 Θ_V 又包含 V. 稳定性的一般定义为：称 $F : D \to E$ 的某个解 $\bar{e} = F(\bar{d})$ 相对于邻域集 Θ_D, Θ_E 为稳定的，是指对任何 $\alpha \in N(\bar{e})$，均可找到 $\beta \in N(\bar{d})$，使得只要 $d \in \beta$，就有

$$F(d) \in \alpha.$$

上述一般定义可以具体化如下：

1. 若 D 为初态集 C_0，E 为输出 Y，F 为对某给定输入 $\hat{x}$ 的初值响应函数 ρ_0，则该定义给出特定响应 $\hat{y} = \rho_0(C_0, \hat{x})$ 对输入、输出变化的邻域集 Θ_{C_0}, Θ_y 的稳定性.

2. 当 D 为输入集 X，其邻域集定义为 Θ_x，则该定义给出特定响应 $\hat{y} = \rho_0(C_0, \hat{x})$ 对输入、输出变化的邻域集 Θ_x, Θ_y 的稳定性，即输入-输出稳定性.

3. 设系统 $S \subset X \times Y$ 中含有若干参数 $d \in D$，系统行为可分为若干模式 $e \in E$，当要求系统运动必须遵守某种模式 e^* 时，可定义 $N(e) = \{e^*\}$(单元素集合)，这时系统行动模式 $e^* \in E$ 为结构稳定(亦称对参数 d 具有鲁棒性)，是指对每个满足 $\hat{e} = F(\hat{d})$ 的 $\hat{d} \in D$，均存在 $\beta \in N(d)$，使只要 $d \in \beta$，就有

$$F(d) = \hat{e}.$$

对种种不同类型的稳定性概念，很难有一般的判据. 具有一定普遍意义的是李亚普诺夫型定理，它建立在广义距离函数和函数值的“序”概念的基础上，亦有许多相应的、实用的表达形式.

系统响应稳定性(stability of system response) 系统输出响应 Y 对于初始状态 C_0 变化的稳定性. 在一般稳定性定义中，当输入 $\hat{x}$ 给定时，集 D 为初态 C_0，集 E 为输出 Y，而函数 F 则为系统初值响应函数 ρ_0. 人们可以得到定义的具体形式为：设 Θ_C 和 Θ_Y 分别为给定的 C 和 Y 的子集族，则系统响应

$$\hat{y} = \rho_0(C_0, \hat{x})$$

称为对于 Θ_C, Θ_Y 和给定输入 $\hat{x}$ 为稳定的，当且仅当

$$(\forall \alpha \in N(\hat{y}))(\exists \beta \in N(C_0))$$
$$(\forall C_0)(C_0 \in \beta \to \rho_0(C_0, \hat{x}) \in \alpha).$$

其中 $N(\hat{y})$ 和 $N(C_0)$ 分别为 $\hat{y}$ 和 C_0 的邻域族.

汉克尔矩阵(Hankel matrix) 可完整描述线性系统的一种参数矩阵形式. 对 m 输入、p 输出线性系统的马尔可夫序列 $\mathscr{A} = \{\mathscr{A}_i, i = 1, 2, \cdots\}$，设 s, t 为任意正整数，则 $\mathscr{A}$ 的 s, t 阶汉克尔矩阵定义为 $ps \times mt$ 阶分块矩阵，即

$$\mathscr{H}_{s,t}(\mathscr{A}) \triangleq \begin{bmatrix} \mathscr{A}_1 & \mathscr{A}_2 & \cdots & \mathscr{A}_t \\ \mathscr{A}_2 & \mathscr{A}_3 & \cdots & \mathscr{A}_{t+1} \\ \vdots & \vdots & & \vdots \\ \mathscr{A}_s & \mathscr{A}_{s+1} & \cdots & \mathscr{A}_{s+t-1} \end{bmatrix}.$$

显然，汉克尔矩阵给出马尔可夫序列的内容，是线性系统的又一种完整描述. 若已知系统的 A, B, C 矩阵，构造其能达性矩阵

$$R_k(A, B) = [B, AB, \cdots, A^{k-1}B],$$

和能观性矩阵

$$O_k(A, C) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{bmatrix},$$

则它们与汉克尔矩阵之间的关系可归结为：三元组 (A, B, C) 是马尔可夫序列 $\mathscr{A}$ 之实现的充分必要条件为：对所有 s 和 t，均有

$$O_s(A, C)R_t(A, B) = \mathscr{H}_{s,t}(\mathscr{A}).$$

阿贝尔群机器(Abelian group machine) 一种建立在阿贝尔群上的形式系统. 这时系统可定义为

$$\Sigma = (U, X, Y, \delta, \beta),$$

式中 U, X, Y 都是阿贝尔群，且有三个同态

$$F: X \to X,\ G: U \to X,\ H: X \to Y,$$

使对所有 $x \in X$ 和 $u \in U$，均有

$$\delta(x, u) = Fx + Gu,\ \beta(x) = Hx.$$

将 R 模中之 R 改为整数 Z，R 模就成为阿贝尔群，R 线性映射即为阿贝尔群同态.

范畴(category) 一种代数对象 $\mathscr{K}$. 它包括：一类个体，对任何个体之序偶 (A, B) 均有集 $\mathscr{K}(A, B)$(可能是空集)，其中元素为由 A 到 B 的 $\mathscr{K}$ 映射，亦即 $f \in \mathscr{K}(A, B)$，$A \xrightarrow{f} B$ 或 $f: A \to B$；对个体的任何有序三元组 (A, B, C)，均有复合映射：

$$\mathscr{K}(A, B) \times \mathscr{K}(B, C) \to \mathscr{K}(A, C):$$
$$(A \xrightarrow{f} B, B \xrightarrow{g} C) \mapsto A \xrightarrow{gf} C,$$

且这种 $\mathscr{K}$ 映射的合成满足结合律，并存在恒同映射.

范畴的例子有：所有集合组成的类 $\mathscr{S}$，所有幺半群类 Mon 组成的幺半群范畴，所有个体 R 模组

成的 R-Mod 等. 比较直观地说,范畴是有结构的集,连同保存结构的映射的集合.

因果算子(causal operator)　一类输入-输出算子. 即反映输入与输出间因果先后关系的算子. 将输入-输出系统定义为希尔伯特空间 H 中输入、输出关系的算子时,需选用满足物理可实现性即因果性条件的那部分算子. 这些算子应保证输入函数在任何时刻 t 的值只影响输出函数在 t 之后的变化. 例如用

$$y(t)=\int_{-\infty}^{\infty}h(t,u(\tau))\mathrm{d}\tau$$

描述输入 $u(\cdot)$ 和输出 $y(\cdot)$ 的关系时,因果性条件要求当 $t<\tau$ 时 $h(t,u(\tau))=0$. 在线性系统情形,因果算子可一般地表示为

$$y(t)=y(0)+\int_0^t h(t,\tau)u(\tau)\mathrm{d}\tau.$$

不考虑因果性进行系统设计时,会导致物理不可实现或不稳定的结果. 形式化地,若用 $\varepsilon=\{E^\tau:\tau\in[0,\infty)\}$ 表示希尔伯特空间 H 上的一族正交投影. 满足: $E^{\tau_1}\supset E^{\tau_2}$, 若 $\tau_1\geqslant\tau_2$; $E^0=0$(零算子), $E^\infty=I$(恒同算子); 且 E 在强极限意义下是封闭的,则 (H,ε) 上的算子 A 称为因果的,若对所有 $E^\tau\in\varepsilon$ 和任何 x, $y\in H$ 均可由等式 $E^\tau x=E^\tau y$ 推出 $E^\tau Ax=E^\tau Ay$.

因果算子的基本性质有:

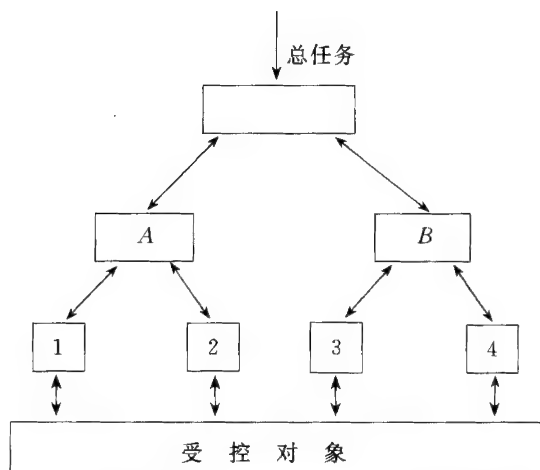
1. $E^\tau A=E^\tau AE^\tau$.
2. $AE_\tau=E_\tau AE_\tau(E_\tau\triangleq I-E^\tau)$.
3. 子空间 $H_\tau=E_\tau(H)=(I-E^\tau)H$ 在 A 之下不变等.

反因果算子(anticausal operator)　一种输入-输出算子. 指时间逆序时的因果算子. 记 $\mathscr{E}^*$ 为与 $\mathscr{E}$ 互为逆序的分解结构,若它在 $(H,\mathscr{E}^*)$ 上是因果的,则 $(H,\mathscr{E})$ 上的算子 A 称为反因果的.

狭义因果算子(strictly causal operator)　一种因果算子. 即具更强因果性的一类算子. 其中,任一算子 A 对每个 $\varepsilon>0$,均存在 ε 的划分 $\mathscr{P}'$,使得对每个比 $\mathscr{P}'$ 更为精细的划分 $\mathscr{P}$ 中的各正交投影 E^i,均有 $\|\Delta^iA\Delta^i\|<\varepsilon$,其中 Δ^i 为对应每区间 $(\tau_{i-1},\tau_i]$ 的正实数.

无记忆算子(memoryless operator)　一种简单算子. 基本特征是输出函数在 t 时刻的值 $y(t)$ 只依赖于其同时刻输入值 $u(t)$. 它可用一简单的静态函数关系 $y(t)=F[u(t)]$ 来描述. 有时它也包括死区、滞环等非单值函数的情形,实际应用中常可遇到.

多级递阶系统(multilevel hierarchical system)　简称递阶系统. 一类具有多级递阶控制结构的大系统. 大规模系统可分解为处在不同层次上的众多子系统,同级子系统之间没有横向联系. 每个子系统只有一个上级,但可能管理若干个下级,其下级个数由于人类管理能力的限制,一般也不超过 10 个. 传统的社会或企业管理结构多为这种形式. 这在一定技术条件下是合理的,但假设系统中信息只能纵向传送. 具有一定局限性. 由于信息技术迅速发展,信息处理能力大为提高,在社会组织和企业中都可能增加各部门的"横向"联系,形成了"多级异阶系统"的新型结构.



多级递阶系统

递阶系统(hierarchical system)　见"多级递阶系统".

多级异阶系统(multilevel hierarchical system)　见"多级递阶系统".

开放系统(open system)　对外界开放的一类系统. 具体常指与外界环境有物质、能量及信息交换的系统,这种交换可导致系统演化,并形成种种有序结构. 许多物理、化学系统、生命系统、地球以及社会、经济系统等均属这一范畴,是系统理论研究的主要对象.

简单系统(simple system)　规模不很大且结构简单的系统. 钱学森系统分类中的一类,其共同特点是元素或子系统的种类较少,关系单纯,层次少,由元素或子系统的描述经过直接综合即可得到系统整体行为的描述,亦即可以用还原论方法来研究的系统. 经典控制论和大系统控制论处理的均为简单系统.

简单巨系统(simple giant system)　规模巨大但结构简单的系统. 钱学森系统分类中的一类,特点是系统规模巨大(元素或子系统的个数超出常规处理能力),但元素或子系统种类很少,相互关系简单,通常只有微观和宏观两个层次,通过统计综合即可从微观描述过渡到对系统宏观整体的描述. 热力学、统计力学、耗散结构理论、协同学讨论的都是简单巨系统.

复杂巨系统(complex giant system)　规模巨

大且结构复杂的系统. 钱学森系统分类中的一类，其特点是系统不仅规模巨大，属巨系统范畴，而且元素或子系统种类繁多，本质各异，相互关系复杂多变，存在多重宏、微观层次，不同层次之间关联复杂，作用机制不清，因而不可能通过简单的统计综合方法从微观描述推断其宏观行为. 人脑、人体、生物、生态、地理环境以及许多宇宙现象等都属于复杂巨系统，其中社会系统是特殊复杂巨系统.

开放的复杂巨系统理论(open complex giant system theory)　钱学森倡导的系统研究的新领域. 研究的背景仍然是人脑、人体、生物、生态、社会、地理环境等复杂巨系统，并强调其对环境的开放性. 这里的"开放性"主要强调系统与环境的信息交流及由此获得的演化、学习和适应能力. 系统通过获取信息、积累经验，具有一定的"预见性"，能够主动地适应环境并改进自己的决策行为. 这类系统包括全球、海洋、生态、社会经济等与人类生存发展有密切关系的复杂原系统，有明确的背景和实际意义.

从定性到定量综合集成法(meta-synthesis from qualitative to quantitative)　处理开放的复杂巨系统问题的基本方法. 基于知识与经验、人与计算机相结合研究开放的复杂巨系统的方法，也是处理开放的复杂巨系统预测和决策问题的具有可操作性的方法. 开放的复杂巨系统目前尚属有待发展、完善的科学领域. 它既不能直接应用从微观描述过渡到宏观描述的统计方法，更不能像简单系统那样由组分的描述直接综合出其整体的性态. 可行的办法是把各方面专家的实践经验、分门别类的科学知识和大量的统计数据资料综合起来，把人脑的智慧和计算机处理信息的能力结合起来，发挥各自的优势，把专家们局部的、零散的定性的经验知识上升为关于系统整体的定量认识. 为把理论知识、专家经验和统计数据综合起来，需要一种智力框架即数学模型，需要定量计算. 模型的建立本身是认识与实践循环往复的过程. 通过反复演示，不断收集专家的经验性判断和感受，反复检验和修改模型，直到仿真演算的系统行为特性得到专家们认可为止. 这样得到的模型是对开放的复杂巨系统可以信赖的理论描述，分析求解模型即可获得关于系统整体的定量认识.

人工生命(artificial life)　一门利用计算机模拟和研究各种类似生命现象所具有之共同特征的新学科. 兰顿(Langton, C. G.)提出的一种复杂适应系统模型，试图通过对可能生命形式(类生命)的研究来加深对现实的地球生命形式的理解. 人工生命理论认为：生命是物质自组织所产生的性质，而不是物质自身固有的属性. 与传统生物学不同，人工生命关注的是生命的形式基础，把有机体看做简单单元构成的大群体，由下到上综合地进行研究. 通过简单规则支配的、彼此非线性相互作用的对象的大量聚集来建构类生命的整体动力学. 人工生命的关键概念是涌现行为，把自然生命看做由大量无生命分子有组织地相互作用而涌现出来的. 在这个过程中，不存在对于组分行为的全局控制者. 相反，每个局部的行为是自主的，生命是从所有个体行为的局部相互作用中涌现出来的动力学行为.

有势系统(system with potential)　一类由势函数导出的有势力来决定系统运动方程的系统. 力学系统的势就是机械势能，称为力学势. 热力学系统的自由能是它的势，称为热力学势. 化学反应系统的势指化学势或化学亲和势. 势函数决定了系统演化行为的走向. 它可表示为其状态变量的函数，有时还取决于反映环境对系统的影响和制约的控制参数. 设系统状态变量为 $X=(x_1, x_2, \cdots, x_n)$，又有 m 个控制参数向量 $C=(c_1, c_2, \cdots, c_m)$，其势函数的一般形式为 $V(X, C)=V(x_1, x_2, \cdots x_n; c_1, c_2, \cdots, c_m)$. 例如，弹簧势能 E 与状态变量(弹簧拉长量)x 的关系为 $E=-kx^2$，这就是弹簧的势函数表达式. 其中弹性系数 k 可作为控制参数. 人们对有势系统的结构、性能和演化行为的研究，都可以归结为对势函数的研究. 这是有势系统突出的特点. 若势函数 $V(X, C)$ 足够光滑，它对状态变量的一阶导数称为梯度. 在力学中，势函数的负梯度

$$f_i=-\frac{\partial V(X, C)}{\partial x_i}$$

即代表势函数确定的作用力在 x_i 方向上的分量，有重要物理意义. 有势系统又称梯度系统，初等突变理论也是关于梯度系统奇异性的理论.

黑塞矩阵(Hessian matrix)　一种特殊矩阵. 是以多元函数的二阶偏导数为元素构成的矩阵. 它是以数学家黑塞(Hesse, L. O.)命名的矩阵. n 元光滑函数 $f(x_1, x_2, \cdots, x_n)$ 的黑塞矩阵，记为 Hf，定义为

$$Hf=\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

自组织系统理论(self-organizing system)　复杂系统理论的分支之一. 主要研究系统组织化结构自动形成的机制和过程. 其具体含义有二：其一指耗散结构理论与协同学等所研究的开放系统，由其与外界进行物质和能量交换导致某种有序结构的自组织过程；另一含义则为人类设计的能依环境和功能要求的变化自动辨识和判断局势搜索新的有序结构，调整和重组具有更为灵活而复杂的功能的系统.

这类系统大多与高新技术、社会经济等复杂问题有密切联系.

拟阵(matroid) 亦称矩阵胚.向量独立性概念的一种推广.由美国数学家惠特尼(Whitney,H.)于1935年提出.其定义如下:如果E为有限集,$\mathscr{S}$为E的某些子集的集合,并满足条件:

1. 空集$\varnothing\in\mathscr{S}$.
2. 若$Y\subseteq X\in\mathscr{S}$,则$Y\in\mathscr{S}$.
3. 若$X,Y\in\mathscr{S}$,$|X|=|Y|+1$,则存在$x\in X$,$x\overline{\in}Y$,使$Y\cup\{x\}\in\mathscr{S}$(这里$|X|$表示X中元素个数).

则称$(E,\mathscr{S})$为一个拟阵.

满足上述条件1和2的$(E,\mathscr{S})$称为一个独立系.拟阵首先在图论研究中引起注意,它使图论中某些概念易于理解,产生了许多新的结果,并使原有若干证明大为简化.在组合优化中也有重要应用,特别是三个以上拟阵交的独立系统问题至今仍是难题.它的研究对于最短路问题、最小树形图问题、旅行商问题等均有重要意义.

矩阵胚(matroid) 即“拟阵”.

马尔可夫序列(Markov sequence) 可完整描述线性系统的一种参数序列形式.由线性系统$\dot{x}=Ax+Bu,y=Cx$或$x_{k+1}=Ax_k+Bu_k,y_k=Cx_k$的系数矩阵构成的序列.$\mathscr{A}=\{\mathscr{A}_i,i=1,2,\cdots\}$,其中$\mathscr{A}_i=CA^{i-1}B,i>0$称为马尔可夫序列.而各$\mathscr{A}_i$则称为马尔可夫参数.马尔可夫序列也是线性系统的一种完整描述,若找到矩阵A,B,C,使有$A_i=CA^{i-1}B$的分解形式,则(A,B,C)就是$\mathscr{A}$的一个实现.

R模(R-module) 向量空间概念的一种推广.用以描述更一般的线性系统.在有单位元的环R和具二元运算$(a_1+a_2)=a_3(a_1,a_2,a_3\in A)$的集$A$之间,若定义了满足下列条件的映射$A\to A:a\mapsto r\cdot a$,则称$A$为一个$R$模:

1. A对“+”运算是个阿贝尔群.
2. $1\cdot a=a$.
3. $(r_1r_2a)=r_1(r_2a)$.
4. $(r_1+r_2)a=r_1a+r_2a$.
5. $r(a_1+a_2)=ra_1+ra_2$.

由定义可见,若R为域,则R模正是R上的向量空间.利用R模可以描述最一般的离散时间定常线性系统$\Sigma=(U,X,Y,\delta,\beta)$或$(X,F,U,G,Y,H)$,其中输入空间$U$、状态空间$X$和输出空间$Y$均为$R$模,$\delta(x,y)=Fx+Gu,\beta(x)=Hx$表示输入、状态迁移和输出映射.

$R[z]$模($R[z]$-module) R模概念的一种推广.给定有单位元的环R,定义R^{ξ}为R中元素的无穷序列.将R^{ξ}中元素$(\cdots u_k\cdots u_1u_0)$形式地表示为多项式$\cdots+u_kz^k+\cdots+u_1z+u_0$,其集合记为$R[z]$.在$R[z]$上可形式地定义多项式加法和乘法,显然它们分别表示$R^{\xi}$中序列的逐个相加和“卷积”.类似于$R$模(参见“$R$模”),可以定义系统状态空间$X$上的$R[z]$模.对线性系统$(X,F,U,G,Y,H)$,易证

$$\Big(\sum_i a_iz^i\Big)x=\sum_i a_iF^ix,$$

其中

$$\sum_i a_iz^i\in R[z].$$

利用$R[z]$模,卡尔曼(Kalman,R.)等人对连续或离散时间线性系统的实现问题给出了一种统一的、形式化的处理,并给出了相应的算法.

伏泰拉级数展开系统模型(Volterra series expansion system model) 一类系统模型.线性系统的脉冲响应函数卷积模型在非线性情形的一种推广.其一般形式的系统输出可表示为

$$y(t)=W_0(t)+\int_0^t W_1(t,s)u(s)\mathrm{d}s$$
$$+\int_0^t\int_0^{s_1}W_2(t,s_1,s_2)u(s_2)u(s_1)\mathrm{d}s_2\mathrm{d}s_1+\cdots,$$

其中,$W_0,W_1,W_2,\cdots$诸函数分别称为各阶的伏泰拉核.

对所谓线性解析系统$\dot{x}=f(x,u)\triangleq f(x)+u(t)g(x)$,$y=h(x)$,$x(0)=x_0$,这里$f,g,h$均为解析向量场,只要$\dot{x}=f(x)$在$[0,T]$上具有满足$x(0)=x_0$的解,则该系统的输入、输出行为具有惟一的伏泰拉级数展开表达式.

若对双线性系统$f(x)=Fx,g(x)=Gx,h(x)=x$,u为标量输入,则其伏泰拉核为

$$W_n(t,s_1,s_2,\cdots,s_n)$$
$$=\mathrm{e}^{Ft}\mathrm{e}^{-Fs_1}G\mathrm{e}^{Fs_1}\mathrm{e}^{-Fs_2}G\mathrm{e}^{Fs_2}\cdots\mathrm{e}^{-Fs_n}G\mathrm{e}^{Fs_n}x_0,$$

且可证明该展开式当u为局部有界时是全局收敛的.

对伏泰拉核进行多维拉普拉斯变换,还可导出一种非线性传递函数的概念.

动 态 系 统

动态系统(dynamic system) 描述事物动态演化的数学模型.最初由n维空间R^n中自治微分方程

$$\dot{x}=f(x),\ x\in\mathrm{R}^n,\ f:\mathrm{R}^n\to\mathrm{R}^n$$

所定义,当满足某些适定性条件时,其以$x(0)=x_0$为初值的解$\alpha(t,x_0)$满足条件:

1. $\alpha(0,x)=x$.
2. $\alpha(s+t,x)=\alpha(s,\alpha(t,x))$.
3. $\alpha(t,x)$对(t,x)连续.

后来则直接把满足条件1至3的映射

$$\alpha: \mathrm{R} \times \mathrm{R}^n \to \mathrm{R}^n$$

定义为 R^n 中的动态系统，或称为方程 $\dot{x}=f(x)$ 的流. 点集 $O_\alpha(x)=\{\alpha(t,x)\mid t\in\mathrm{R}\}$ 则称为相轨道.

在上述定义中，还可将 R^n 推广为某紧致 C^∞ 微分流形 M，得到流形 M 上的动态系统；或将实数时间 $t\in\mathrm{R}$ 改为整数的 $t\in\mathrm{Z}$，而得到离散(时间)动力系统.

一般系统的动态表示(dynamical representation of general system) 用状态迁移模型来表示给定的一般系统输入-输出映射的问题. 设给定一般的时间系统 $S\subset A^T\times B^T$，且它可用函数族 $\bar{\rho}=\{\rho_t\mid c_t\times X_t\to Y_t; t\in T\}$ 来表示，当且仅当存在函数族 $\bar{\phi}=\{\phi_{tt'}\mid c_t\times X_{tt'}\to c_{t'}\}$，使得 $\bar{\rho}$ 与 S 相容，且 $(\bar{\rho},\bar{\phi})$ 为 S 的一个动态表示.

状态空间(state space) 描述系统运动的一种抽象空间. 指以系统的状态变量为轴所张成的空间，亦即系统所有可能状态的集合. 具体为一般系统定义中能够用于表示时间系统 $S\subset X\times Y$ 演化规则 $\bar{\rho}=\{\rho_t: C\times X_t\to \overline{Y}_t\}$ 和 $\bar{\phi}=\{\phi_{tt'}: C\times X_{tt'}\to C\}$ 的状态 C 的集合. 这时，可称 S 为状态空间 C 中的动态系统.

状态空间的点称为代表点或相点，系统运动轨迹(动力学方程的每个解)亦为状态空间的一个点集合，称为轨线. 独立状态变量个数称为状态空间的维数，通常为非零整数 0,1,2,…特别地，2 维状态空间称为状态平面. 用状态空间描述系统运动，可以把解析问题转化为较直观的几何问题来处理，有时比较方便.

通常希望寻找为实现同一时间系统 S 所需的最小状态空间，这时得到的动态系统描述称为最小实现(参见“最小实现”).

弗罗贝尼乌斯-佩龙定理(Frobenius-Perron theorem) 关于所有元素均为正实数的正矩阵的基本性质的重要定理. 该定理断言：若矩阵 $A>0$，则存在正实数 $\lambda_0>0$ 和 $x_0>0$，使得：

1. $Ax_0=\lambda_0 x_0$.

2. 若 $\lambda\neq\lambda_0$ 是矩阵 A 的任一其他特征值，则 $|\lambda|<\lambda_0$.

3. λ_0 是一个几何重数和代数重数均为 1 的特征值.

梅兹内矩阵(Metzler matrix) 一类实方阵. 按纽曼(Neumann, M.)的定义，梅兹内矩阵应满足条件：$a_{ii}<0$；$a_{ij}\geqslant 0(i\neq j)$. 在阿罗(Arrow)的定义中，只要求满足后一个条件.

闵科夫斯基矩阵(Minkowski matrix) 一类实方阵. 闵科夫斯基矩阵应满足条件：$a_{ij}\leqslant 0(i\neq j)$，且各阶主子式满足 $|a_{ii}|\geqslant 0$.

赫尔维茨矩阵(Hurwitz matrix) 所有本征值均具负实部的矩阵. 其形式化表示为

$$\mathrm{Re}\ \lambda_i(A)<0, \forall i.$$

当线性常系数微分方程的系数矩阵为赫尔维茨矩阵时，该系统是渐近稳定的.

希克斯矩阵(Hicks matrix) 一类方阵. 其特征是所有奇阶主子式均为负，而所有偶阶主子式均为正的矩阵. 其基本性质归结为定理：若 A 为梅兹内矩阵，则 A 为赫尔维茨矩阵稳定的充分必要条件为 A 是希克斯矩阵.

森岛矩阵(Morishima matrix) 一种不可约矩阵. 森岛矩阵的形式可由行列互换而变为

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

其中，$A_{11}\geqslant 0$，$A_{22}\geqslant 0$，均为方形子阵，且 $A_{21}\leqslant 0$，$A_{12}\leqslant 0$.

正线性系统(positive linear system) 一类形如线性的动态系统. 其状态变量始终只能取正实数值. 这种状态空间受限的系统，常用于描述系统包含物理上不可能为负的状态(如人口的年龄分布)的情形，在社会、经济、生态系统研究中常可遇到.

二维系统(two-dimensional system) 具有两维独立“时间”变量的动态系统. 常见于多维数字滤波、图象信息处理、网络综合、分布参数系统等问题. 它的状态空间模型最早是由罗塞尔(Roesser)和乔万尼(Givone)于 1972 年提出的 2D 罗塞尔模型

$$\begin{aligned}
&\begin{bmatrix} x^h(i+1,j) \\ x^v(i,j+1) \end{bmatrix} \\
&= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} x^h(i,j) \\ x^v(i,j) \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u(i,j), \\
&y(i,j) \\
&= \begin{bmatrix} c_1 & c_2 \end{bmatrix}\begin{bmatrix} x^h(i,j) \\ x^v(i,j) \end{bmatrix} + Du(i,j),
\end{aligned}$$

其边界条件为

$$\begin{aligned}
x^h(0,j) &= x^h_{0,j}, \\
x^v(i,0) &= x^v_{i,0} \quad (i,j=0,1,2,\cdots).
\end{aligned}$$

1973 年，阿塔西(Attasi)提出 2D 阿塔西模型

$$\begin{aligned}
x(i+1,j+1) &= A_1x(i+1,j)+A_2x(i,j+1) \\
&\quad -A_1A_2x(i,j)+Bu(i,j), \\
y(i,j) &= Cx(i,j);\ (i,j)\geqslant(0,0).
\end{aligned}$$

1976 年，富纳西尼(Fornasini)和马舍西尼(Marchesini)提出了第一富纳西尼-马舍西尼模型

$$\begin{aligned}
x(i+1,j+1) &= A_0x(i,j)+A_1x(i+1,j) \\
&\quad +A_2x(i,j+1)+Bu(i,j), \\
y(i,j) &= Cx(i,j);\ (i,j)\geqslant(0,0)
\end{aligned}$$

和第二富纳西尼-马舍西尼模型

$$x(i+1,j+1)=A_0x(i,j)+A_1x(i+1,j)$$

$$+A_2x(i,j+1)+B_0u(i,j)$$
$$+B_1u(i+1,j)$$
$$+B_2u(i,j+1),$$
$$y(i,j)=Cx(i,j)+Du(i,j);(i,j)\geqslant(0,0).$$

1985 年,库列克(Kureck)提出了线性数字系统的 2D一般模型

$$x(i+1,j+1)=A_0x(i,j)+A_1x(i+1,j)$$
$$+A_2x(i,j+1)+B_0u(i,j)$$
$$+B_1u(i+1,j)$$
$$+B_2u(i,j+1),$$
$$y(i,j)=Cx(i,j)+Du(i,j);\ (i,j)\geqslant(0,0).$$

借用常规线性系统的概念和方法,对二维系统也获得了稳定性、能控性、能观性、状态反馈、极点配置、解耦、最优控制、参数辨识等一系列结果.这些结果也可推广到广义二维系统或多维系统的情形.二维系统及更一般的多维系统在数字滤波、图象处理、网络综合、地质勘探和地震分析、气象预报等方面均有重要的应用.

广义系统(generalized system) 亦称奇异系统或描述器系统.一类蜕化的或推广的系统模型.线性系统各状态变量的微分之间不独立的情形,实质上是若干代数式和微分式的耦合.其连续时间情形的数学模型为 $E\dot{x}=Ax+Bu, Y=Cx$,而 E 为一不满秩的矩阵.这类系统最早于 1974 年提出,其背景为某些电气网络、经济学中的列昂节夫(Leontief, W.)的投入产出模型、哈罗德(Harrod)和萨缪尔森(Samuelson, P. A.)的国民经济模型,以及若干人口、生态问题等.它也可能是奇异摄动系统中令小参数 $\varepsilon=0$ 时的一级近似.将状态变量 x 重新组合相当于矩阵 E 进行相似变换,可以把系统变为 m 阶($m=\operatorname{rank}E$)微分方程和 $n-m$ 个线性代数式的组合.当$\det(\lambda E-A)$不恒为 0 时,广义系统称为可解的.广义系统除了对应于常规极点的指数运动模式,还具有无穷远极点对应的脉冲模式.它可能包含输入的导数项,从而不满足因果性条件.广义系统的研究成果包括:用矩阵束理论导出的可解性、能控能观性、极点配置、最优控制,以及随机广义系统基于观测与估计的动态反馈等问题.

奇异系统(singular system) 即"广义系统".

描述器系统(descriptor system) 见"广义系统".

多时标系统(Multi-time-scale system) 存在多种不同时间尺度的系统.其实际背景包括如电力系统、细长物体振动系统,以及由慢变量支配快变量的协同学系统等.设系统状态变量依变化快慢不同可分为 x,y,z 三组,则系统微分方程可写成如下形式:

$$\dot{x}=F(x,y,z);$$
$$\varepsilon\dot{y}=G(x,y,z);$$
$$\mu\dot{z}=H(x,y,z);$$

这里 $1\gg\varepsilon\gg\mu$.当方程两边同乘以小参数 ε 或 μ 时,可见 x 的变化率较 y 和 z 小很多,在研究 y 和 z 的变化时,可近似看成常数,从而可以当做广义系统来处理(参见"广义系统").又因令摄动参数 ε,μ 趋于零时,系统的阶数将降低,它也是一类奇异摄动系统(参见"奇异摄动系统").相应地各种处理方法也都可以用于多时标系统.

奇异摄动系统(singularly perturbed system) 由小参数摄动导致奇异性的系统.对含有摄动参数 ε 的系统问题族 P_ε,若它的解能表示为 ε 的幂级数

$$y_\varepsilon\approx y_0(x)+\sum_{i=1}^{\infty}y_i(x)\cdot\varepsilon^i,$$

$x\in D$,而且当 $\varepsilon\to0$ 时该式对 $x\in\overline{D}$ 是一致收敛的,则称 P_ε 为正则摄动问题,否则即为奇异摄动问题.在电力系统、细长杆振动、具寄生小参数的电气网络、粘性流体边界层等问题中均可遇到奇异摄动.其典型数学模型为

$$\dot{x}(t,\varepsilon)=F(x,z,t,\varepsilon),$$
$$\varepsilon\dot{z}(t,\varepsilon)=G(x,z,t,\varepsilon),$$

或较简单的线性系统

$$\dot{x}(t,\varepsilon)=Ax+Bz,$$
$$\varepsilon\dot{z}(t,\varepsilon)=Cx+Dz,$$

其中 ε 为小参数,x 为慢变分量,z 为快变分量.处理奇异摄动系统的方法主要是边界层法,可用于数值求解或处理相应的稳定性和反馈控制问题.把快、慢变模式分开,可以把系统化成低阶模型以简化计算,也可克服数值积分步长不一致带来的困难.

隐式系统(implicitive system) 一种微分动力系统.指用隐函数形式给出的系统.其一般形式为

$$F(\dot{x},x,u,t)=0,$$

其中 $\dot{x},x\in\mathbb{R}^n, u\in\mathbb{R}^m, F:\mathbb{R}^n\times\mathbb{R}^m\times T\to\mathbb{R}^n$.当$\frac{\partial F}{\partial\dot{x}}$满秩,从而在某区域中可单值地解出 $\dot{x}=f(x,u,t)$时,它退化为常规的系统.而在

$$\det\left(\frac{\partial F}{\partial\dot{x}}\right)=0$$

附近可能具有某些奇异性,须作进一步的分析.

微分代数系统(differential algebraic system) 一种隐式系统.即由微分方程 $\dot{x}=f(x,t)$和代数方程 $g(\dot{x},x,t)=0$ 耦合而成的系统.当系统线性时,它可写成 $E\dot{x}=Ax+Bu$ 形式,即为广义系统或描述器系统,亦可看成是奇异摄动系统当 $\varepsilon\to0$ 时的极限形式.

集动态系统(set dynamic system) 一类动态系统模型.用最一般的集合论语言描述输入 I 和输

出 O 之间具时间参数集 T 的变换关系 H 的系统. 以离散时间情形 $T=\{0,1,2,\cdots\}$ 为例,设输入为时间序列 $\{u(0),u(1),\cdots\}$ 的集合 I,其中 $u(i)\in U$,系统状态序列集为 $\{x(0),x(1),\cdots\}$ 的集合 S,其中 $x(i)\in X$,以及系统输出序列集为 $\{y(0),y(1),\cdots\}$ 的集合 O,其中 $y(i)\in Y$,则集动态系统可表示为六元组

$$\mathrm{SDS}=\{U,Y,X,T,f,g\}.$$

其中 $f:X\times U\to X$, $g:X\times U\to Y$ 分别表示状态迁移函数 $x_{k+1}=f(x_k,u_k)$ 和输出观测函数 $y_k=g(x_k,u_k)$. 这样,给定一个输入序列和一个初始状态,就可逐步推出整个输出序列,亦即有映射 $H:I\to O$,称其为全局映射. 在此意义下集动态系统也可用三元组 (I,O,H) 来表示. 当 U,X,Y 均为实(或复)数域时,它们就是熟知的线性或非线性系统. 集动态系统可经互联运算组成更复杂的大系统,常见的有串联、并联和反馈等形式. 这里要求各子系统元素之间具有兼容性,并对相关的运算给出定义. 具不同六元组描述的集动态系统可能具有相同的三元组形式,即它们的映射 H 可能相同. 这时称这些六元组系统为等价的. 它们共同的 (I,O,H) 系统就是由此等价类定义的商系统集合. 与此相关的也要研究系统的能控性、能观性、最小实现等问题. 集动态系统可定义在任意集合上,具有极广泛的代表性. 它不但适用于通常由连续变量描述的物理系统,也可用来研究离散、模糊、心理行为等更一般的系统.

李亚普诺夫稳定性(Liapunov stability) 一种稳定性概念. 由俄国李亚普诺夫(Ляпунов, A. M.)于1892年提出和研究. 它最初源于常微分方程系统,而后被推广到一般动态系统. 其含义为系统长远运动状态对于初始状态扰动的连续依赖性. 设动态系统输入 $\hat{x}$ 为零或固定,由 $\phi_{0t}:C\to C$ 规定状态迁移关系. $\hat{c}$ 为平衡点,亦即 $\phi_{0t}(\hat{c})=\hat{c},\forall t$. 称平衡点 $\hat{c}$ 为李亚普诺夫稳定的,若有

$$(\forall\varepsilon)(\exists\delta)(\forall c)(\|c-\hat{c}\|<\delta \to(\forall t))(\|\phi_{0t}(c)-\hat{c}\|<\varepsilon),$$

亦即:任给 ε 均存在 δ,使对任何与 $\hat{c}$ 距离小于 δ 的状态 c,均保证在其后任何 t 时刻与 $\hat{c}$ 的距离都保持小于 ε. 进而还可定义相对于不变集合和任意邻域族的李亚普诺夫稳定性:称集 $C'\subset C$ 为相对于状态迁移关系 ϕ 和邻域族 θ 为稳定的,当且仅当

$$(\forall\alpha\in N(\phi(c')))(\exists\beta\in N(c'))(\phi(\beta)\subset\alpha),$$

亦即:任给 $\phi(c')$ 的邻域族中某邻域 α,均存在 c' 的邻域 β,使得由 β 中演化出来的轨线均在 α 中.

孤立轨道稳定性(stability of an isolated trajectory) 对某给定状态迁移关系族 $\phi(\hat{c})$ 的李亚普诺夫稳定性. 亦即特定轨道 $\phi(\hat{c})$ 对其自身运动的回归性. 一个典型例子是泊松稳定性. 设具有固定输入 $\hat{x}$ 的动态系统 $\hat{z}_c(t)=\phi_{0t}(c,\hat{x}^t)$,称轨道 $\hat{z}_c:T\to c$ 为正泊松稳定是指:存在 $\hat{t}$,使对任何 $\alpha\in N(\hat{z}_c(\hat{t}))$ 和任何 t,均可找到 t' 使得只要 $t\leqslant t'$ 就有 $\hat{z}_c(t')\in\alpha$.

半动态系统(semi-dynamic system) 定义于拓扑空间的一般动态系统. 把动态系统视为一般只具一定弱拓扑结构的相空间中的满足一定条件的相映射. 其定义如下:设 X 为豪斯多夫拓扑空间(相空间),则二元对 $[X,\pi]$ 称为半动态系统,若 $x\in X$, $T=\mathrm{R}^+$ 为时间集,且映射 $\pi:X\times T\to X$ 满足条件:

1. 初值条件 $\pi(x,0)=x$.
2. 半群条件 $\pi(\pi(x,t),s)=\pi(x,t+s)$.
3. 连续性条件,即 π 在乘积空间 $X\times T$ 中连续.

当时间集 $T=\mathrm{R}$ 时,上述即为(拓扑)动态系统的定义. 其本质区别在于:后者中的 $s,t\in\mathrm{R}$,允许系统沿时间逆向演化;而前者 $s,t\in\mathrm{R}^+$,系统只能正向演化,即其过程是不可逆的,其"过去行为"可能是无定义的或不确定的. 半动态系统的例子可以是非自治常微分方程,具有时滞的常微或偏微分方程、随机微分方程等. 当上述的时间集 T 改为非负整数集 Z^+ 时,所定义的即为离散半动态系统.

拓扑动态系统(topological dynamic system) 见"半动态系统".

时延系统(time-delay system) 具时间延迟的动态系统. 系统在 t 时刻的行为与其状态在某一时段 τ 之前的输入或状态值有关的系统. 简单的典型例子如

$$\dot{x}(t)=F[x(t),x(t-\tau)].$$

注意,在区间 $[0,T]$ 求解这类方程时,其初值需给定为整个区间 $[-\tau,0]$ 上的状态值. 由于信息滞后,时延系统的镇定和控制均有许多特殊的困难.

微分包含系统(differential inclusion system) 一种动态系统. 描述具不确定演化规则的系统模型. 该类系统于20世纪60年代提出. 其演化规则为 $\dot{x}\in g(x)$,而 $g(x)$ 为某一依 x 而变的集合. 它可用于描述具不确定性的动态演化过程. 与传统的微分方程描述的动态系统 $\dot{x}=f(x,t)$ 不同,微分包含系统的描述形式为

$$\dot{x}\in F(x,t),$$

其中 F 为与系统状态 x 和时间 t 有关的集值映射,由于微分包含系统的演化方向通常是不惟一的,其初值问题的解一般为一轨迹集合. 许多实际问题需要研究微分包含系统,如经济系统、社会系统、生物系统及某些物理系统,以及在未知干扰下的控制系统等. 在这些系统中,存在许多未知因素及未建模部分,使系统动态具有不确定性.

早期的微分包含系统研究主要集中于特定的广义梯度方程和苏联学者菲里波夫(Filipov)的非光滑微分方程. 前者是由梯度方程 $\dot{x}=-\nabla V(x(t))$ 演化

而来，其中 V 是可微势函数；后者源于继电开关控制系统的动力学. 系统地研究微分包含系统始于 20 世纪 70 年代后期. 与其他动态系统的研究一样，不确定性可能表示随机或未知因素，亦可作为设计者选择的范围. 一般的控制系统 $\dot{x}=f(x,u)$ 可看成

$$g(x)=\{f(x,u)\mid u\in U\}$$

时的微分包含式.

微分包含系统的研究内容包括：它的不变集、回归集、吸引集、极限环、可达集的性质和计算，以及最优控制变量 u 的选择. 分析方法主要基于集值映射和存活理论等. 所谓存活理论，是对给定闭和紧致的存活集 $K(t)$，寻求满足 $\dot{x}\in F(x)$ 的解，使得 $x(t)\in K(t)$. 主要理论结果有松弛定理，以及用 ε 网格或分段常值微分包含式的逼近等.

微分动态系统(differential dynamic system) 现代数学的新兴分支之一. 微分动态系统是研究微分流形上连续或离散的可微分动态系统轨道整体(大范围)性质的新兴数学分支. 它的前身可追溯到 19 世纪末，法国科学家庞加莱(Poincaré,(J.-)H.)等人开创的常微分方程定性理论与 20 世纪初期美国数学家伯克霍夫(Birkhoff,G. D.)等人关于拓扑动态系统的工作.

20 世纪 60 年代以来，微分动态系统又有许多重大突破. 它以经典的常微分方程定性理论为背景，以微分拓扑、大范围微分几何以及巴拿赫空间微分学等近代数学分支为工具，对微分流形上常微系统生成的可微流或微分同胚生成的离散动态系统时间轨道整体的拓扑性质或统计性质进行大范围分析，得出许多深刻的结果.

微分动态系统的基本概念可简述如下：设 M 是一 C^r 微分流形，R 为实数集，φ： $\mathsf{R}\times M\to M$ 是 C^r 映射. 如果 φ 满足：

1. $\varphi(0,x)=x$，$\forall x\in M$；

2. $\varphi(s+t,x)=\varphi(s,\varphi(t,x))$，$\forall s,t\in\mathsf{R}$，$x\in M$；

则称 φ 为流形 M 上的 C^r 动态系统(或 C^r 流).

设 f 是 C^r 微分流形 M 上一 C^r 微分同胚，Z 为整数集. 对任何 $i\in\mathsf{Z}$，f 的 i 次迭代记为 f^i，则有：

1. $f^0=id$(恒等映射).

2. $f^{k+l}=f^k\cdot f^l$，$\forall k,l\in\mathsf{Z}$.

于是与流的情形相对应，称流形 M 上的 C^r 微分同胚 f 生成了离散的动态系统 $\{f^i\mid i\in\mathsf{Z}\}$. 流与离散的动态系统是彼此密切相关的研究微分动态系统的两个渠道. 结构稳定性、通有性、遍历性、分支等理论是微分动态系统研究的重要问题.

符号动态系统(symbolic dynamic system) 由符号序列描述的动态系统. 其状态取值于有限符号集合，并按一定规则逐步迭代演化形成的动态系统，其轨道为各种形式的符号串. 一个著名的例子是：将逻辑斯谛映射的[0,1]区间，量化为左、中、右三个符号(L,C,R)，由研究这三个符号构成的序列可以判定系统的定性行为，如解的周期性、混沌的存在性等. 又一个有趣的研究课题则是由各种“语法”规则生成的符号串的形式语言的复杂度和描述能力(参见“符号动力学”).

希尔伯特空间动态系统(Hilbert space dynamic system) 用希尔伯特函数空间中的算子来定义的动态系统模型. 它将输入时间函数变换为输出时间函数，可以利用泛函分析的投影定理等已有结果简捷地得到反馈系统稳定性和最优控制的形式上的解答. 但实现这些结果将遇到因果性导致的物理可实现性的困难，亦成为希尔伯特空间动态系统控制理论要解决的主要问题.

利用赋有时间结构的希尔伯特空间分解在因果性、稳定性、记忆和无源性方面的研究结果，特别是 nest 代数和三角算子模型、因果算子代数等成果，可以对上述问题综合满意的解答，利用赋值于希尔伯特解空间的随机变量处理随机系统估计和控制问题，也得到了一系列有用的结果.

相轨道(phase trajectory) 动态系统术语. 指状态空间中用于表示系统状态变化过程的相点轨迹. R^n 中自治动态系统 α：$\mathsf{R}\times\mathsf{R}^n\to\mathsf{R}^n$ 所确定的点集 $O_\alpha(x)=\{\alpha(t,x)\mid t\in\mathsf{R}\}\subset\mathsf{R}^n$，称为其经由 x 的相轨道(参见“动态系统”).

不动点(fixed point) 亦称平衡点. 相空间的一种单点运动轨道. 在动力系统中有映射的不动点和流的不动点.

一维映射 $x_{n+1}=f(x_n)$，从 x_0 出发，随 n 增大若 x_n 趋向于固定值 x^*，就称它为映射的不动点. 显然不动点满足方程 $x=f(x)$. 同样若 n 很大后，x_n 趋向于两个值 x_1 和 x_2，并在其间来回跳动，则有 $x_2=f(x_1)$，$x_1=f(x_2)$，从而 x_1 和 x_2 满足方程

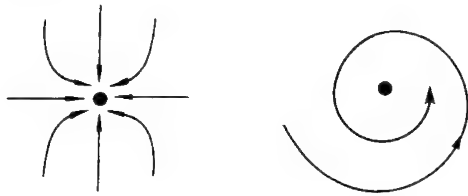
$$x=f(f(x)),$$

从而 x_1，x_2 是映射 $x_{n+1}=f(f(x_n))$ 的不动点，代表映射 $x_{n+1}=f(x_n)$ 的 2 点周期运动.

对连续动力系统 $\dot{x}=f(x)$，当时间 t 很大后，若状态 x 逐渐收敛至 $\bar{x}$，即 $f(\bar{x})=0$，则状态 $\bar{x}$ 就称流的不动点，也称奇异点或临界点.

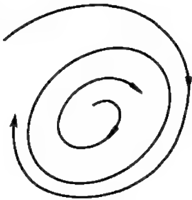
吸引子(attractor) 相空间的一种特殊点集. 在耗散动态系统中，状态随时间演化若可无限趋近于某低维或有限吸收集上，则该吸收集称为吸引子. 常见的吸引子有诸如：点(定常状态)吸引子、极限环(周期状态)吸引子、不变环面(拟周期状态)吸引子和奇怪(混沌)吸引子等. 沿时间增长正向演化的吸引子亦称 ω 吸引子，而沿时间减少反向演化的则称 α 吸引子.

点吸引子(point attractor)　单点形式的吸引子. 由相空间某区域中任一点出发，若系统轨线当 $t \to +\infty$(或 $t \to -\infty$)时，均趋近某个点，则该点称为稳定的(或不稳定的)点吸引子. 常见例子有结点吸引子、焦点吸引子，如图所示. 点吸引子是该系统的不动点，代表定常的运动形态.



结点吸引子　　焦点吸引子

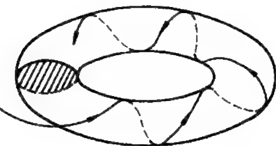
极限环吸引子(limit cycle attractor)　简单封闭曲线的吸引子. 耗散系统在相空间某区域中的轨线，若沿 $t \to +\infty$ 或 $t \to -\infty$ 趋于某孤立闭轨，则该闭轨称为极限环吸引子(如图). 当 $t \to +\infty$ 逼近的称为稳定极限环，$t \to -\infty$ 逼近的则称为不稳定极限环. 电路、机械或自然界中可观察到的许多稳定的周期运动多为极限环的例子.



稳定极限环吸引子

对离散映射，若有一个稳定的不动点，就代表不动点或周期 1 的极限环吸引子. 若有有限数目封闭轨道，则可形成稳定趋近某循环运动的不同周期长度的极限环.

不变环面吸引子(invariant torus attractor)　由环面曲线构成的吸引子. 特指由具不可公度周期的两个周期运动合成的吸引子. 该环面形如轮胎，其上轨线始终不会重复，且其邻域的轨线也都趋向环面，故称为环面吸引子(如图). 环面吸引子表示了一种拟周期的运动形态.



二维环面吸引子

奇怪吸引子(strange attractor)　一种特殊类型的吸引子. 其吸引区中相点运动均向该吸引子逼近，而吸引子内部相点运动是非周期的并具以下基本特征：

1. 该系统局部具正的李亚普诺夫指数，从而其相点至少在相当靠近之后又以指数速度分离.

2. 足够长时间后任意两相点可以任意靠近.

3. 该吸引子集合具分数维结构.

著名的奇怪吸引子有描述气象行为的洛伦兹吸引子、逻辑斯谛迭代映射的混沌吸引子、二维厄农映射的吸引子等(参见“逻辑斯谛映射”、“日本吸引子”、“厄农映射”).

吸引域(basin or domain of attraction)　亦称吸引区. 动态系统术语. 一种与特定吸引集密切相关的集合，该集合中任意点出发的轨道终将趋于该特定吸引集. 其形式定义为：若吸引集 A 的邻域 U 满足条件：对于所有的 $x \in U$ 及所有 $t \geqslant 0$，$\varphi_t(x) \in U$，当 $t \to \infty$ 时，$\varphi_t(x) \to A$，则下述集合构成吸引集 A 的吸引域：

$$\bigcup_{t \geqslant 0} \varphi_t(U).$$

由上述定义可知，每一吸引集具有惟一与之对应的吸引域，且被包含于该吸引域之中. 动力系统理论指出，不同吸引集所对应的吸引域必不相交，且各吸引域被非吸引集的稳定流形隔开.

吸引区(basin or domain of attraction)　即“吸引域”.

拉塞尔不变性原理(Lasalle invariance principle)　关于动态系统当 $t \to \infty$ 时轨道渐近性质的一个命题. 设相空间 X 中包含原点的区域 G 中的标量函数 V 满足：

1. V 在 G 上连续.

2. 对任何 $x \in G$，全导数 $V'(x) \leqslant 0$.

3. G 为正不变的，亦即由 G 出发的任何正向轨道仍在 G 中.

这时，定义 $E = \{y \in \overline{G}: V'(y) = 0\}$，并令 M 为 E 的最大弱不变子集，亦即对任何 $x \in M$ 均存在一条经由 x 的轨道整个处于子集 M 中. 若 $x \in G$ 具有相对的正轨道，则存在 $c \in \mathrm{R}$ 使得

$$x(t) \to M \cap V^{-1}(c).$$

ω 极限点(ω-limit point)　动态系统随时间增长而趋近的极限点. 具体说来，动态系统经过点 x 而令时间 t 增加所得的相轨称为正半轨，并记为 $O_\varphi^+(x)$. 正半轨的极限点称为 x 的 ω 极限点，该极限点集合记为 $\omega(x)$，称为 x 的 ω 极限集.

α 极限点(α-limit point)　动态系统随时间倒退而趋近的极限点. 具体说来，动态系统经过点 x 而令时间 t 减小所得的相轨称为负半轨，并记为 $O_\varphi^-(x)$. 负半轨的极限点称为 x 的 α 极限点，极限点的集合记为 $\alpha(x)$，称为 x 的 α 极限集.

游荡点(wandering point)　动态系统的一种相点. 由该点出发的相轨经足够时间后将不再回归至它的某个邻域. 其形式定义为：对 R^n 中(或流形 M 上)的点 p，若存在它的邻域 $U \subset \mathrm{R}^n$(或 M)和某时间 $N > 0$，使对任意 $t > N$，均有 $\varphi_t(U) \cap U = \varnothing$，即由 U 内出发的轨道均离开 U 而不返回，则称点 p 为游荡点，否则称为非游荡点.

非游荡点(non-wandering point)　见“游荡点”.

大系统稳定性理论(stability theory for large scale system) 通过研究子系统稳定性及其互联关系来判别大系统稳定性的一套概念、方法和结果. 设 n 维大系统 $S:\dot{x}=f(x,t)$ 可以分解为 m 个子系统 $S_i:\dot{z}_i=f_i(z_i,t)+g_i(x,t)$,其中 $z_i\in\mathrm{R}^{n_i}$ 为 S_i 的状态向量,

$$\sum_{i=1}^{m}n_i=n,$$

g_i 为交互关联项. 对这种大规模系统比较保守和可靠的观点是:首先要求各孤立子系统(关联项 $g_i=0$)稳定,例如可用各自的李亚普诺夫函数 $V_i(z_i)$ 的全导数负定加以保证;然后再令关联项足够弱,从而可以不影响由各 V_i 组合而成的全系统李亚普诺夫函数的全导数的负定性,以保证原始大系统仍然稳定.

另一种方法是直接以各 $V_i(z_i)$ 为分量构造向量李亚普诺夫函数(参见"向量李亚普诺夫函数"),再用微分方程的比较原理来判别稳定性. 这些方法均属分解-集结类型,其结果偏于保守,但也反映了大系统安全性的要求.

为保证大系统稳定性,可能还要采用分布式的反馈镇定控制器.

当研究线性大系统稳定性时,通常可限于寻找二次型的李亚普诺夫函数,而正定性之类要求则归结为相应的矩阵应满足的条件.

向量李亚普诺夫函数(vector Liapunov function) 由大系统之各个子系统的李亚普诺夫函数构成的向量函数. 它可用来判别大系统的稳定性. 向量李亚普诺夫函数由美国数学家贝尔曼(Bellman, R.)于 1962 年提出,后于 1966 年由拜利(Bailey, F. N.)将其结合比较原理形成研究大系统稳定性的一种方法. 设由各子系统 S_i 的李亚普诺夫函数 $V_i(z_i)$ 为分量构成向量函数 $V(z)$. 设各 V_i 在孤立子系统中的全导数满足一定的负定性条件,且其间的关联满足一定的约束,通常还要求刻画子系统 V_i 函数衰减性质的几个函数都是同一数量级的,即可保证大系统的全局渐近稳定等性质. 尽管用向量李亚普诺夫函数判别大系统稳定性要比用标量李亚普诺夫函数的条件更多一些、更保守一些,但充分利用大系统的结构特征进行分解,可以改进这一情况.

比较原理(comparison principle) 通过比较微分方程右端直接判断系统稳定性的一种原理. 描述系统动力学特性的微分方程之右端是一个函数,由该函数值的大小可判定方程解的大小,从而可用于判别系统运动的有界性和稳定性. 一般地,对 m 维微分方程 $\dot{y}(t)=W(y(t),t)$,其于 t_0 时过 y_0 点的解 $y(t;y_0,t_0)$ 可能不惟一. 若微分方程右端函数 $W(y,t)$ 对 y 拟单调不减,且 $y_m(t,y_0,t_0)$ 为 t_0 时过 y_0 点的右行最大解,则当连续向量函数 $r(t)$ 满足 $\overline{D}_+r(t)\leqslant W(r(t),t)$, $r(t_0)=y_0$ 时,必有 $r(t)\leqslant y_m(t;y_0,t_0)$. 此命题即为比较原理,$\dot{y}=W(y,t)$ 为比较方程,由它可间接判断 $r(t)$ 的稳定性.

研究大规模复杂系统的稳定性时,可以利用比较原理将之化为一个较为简单的比较系统的稳定性问题,从而得到原系统稳定性的一个较为保守的充分条件.

相空间(phase space) 描述动态系统状态运动的抽象空间. 其各个坐标分别代表系统各状态变量. 质点一维运动的状态可以用它的位置 x 和它的速度 $\dot{x}$ 两个变量来描述,此时相空间 $(x,\dot{x})$ 是二维的,可称为相平面. 同理一般具 n 个质点的系统,其相空间为 $2n$ 维.

对单摆等往复周期运动的系统,其相空间亦可为圆柱面、锥环面等.

无穷维动态系统(infinite dimensional dynamical system) 相空间为无穷维函数空间的动态系统. 最常见的为由偏微分方程所描述的动力系统,如燃烧、热传导、粘性流体运动的纳维-斯托克斯方程,非均匀介质中化学反应的反应扩散方程等,其状态变量(如流体速度、化学成分浓度)实际上是分布于某空间区域的函数,并可以用某无穷维空间区域中的相点来描述,故称为无穷维动态系统. 例如,纳维-斯托克斯方程可以写成下列无穷维动力系统形式:

$$\frac{\mathrm{d}u_k}{\mathrm{d}t}=-\nu k^2u_k+\sum_{i,j}A_{ijk}u_iu_j$$
$$(k=0,1,2,3,\cdots),$$

其中 ν 是粘性系数,右端最后一项是非线性项,A_{ijk} 是系数. 此外,具纯时滞的系统也是无穷维的.

一般动态系统(general dynamic system) 描述不同时刻系统变量之间一般关系的一类模型. 这时除了响应函数,还要引入两个映射族:

1. 输出映射:$\bar{\rho}=\{\rho_t\mid c_t\times X_t\rightarrow Y_t, t\in T\}$.
2. 状态迁移映射:

$$\bar{\phi}=\{\phi_{tt'}\mid c_t\times X_{tt'}\rightarrow c_{t'}, t,t'\in T, t'>t\}.$$

它们还应当满足与系统关系 S 兼容的若干性质.

这一对映射族 $(\bar{\rho},\bar{\phi})$ 即为 S 的动态系统表示,或直接称为动态系统. 注意,并非所有响应族都可以构成动态系统.

守恒系统(conservative system) 亦称保守系统. 具有守恒位势的系统. 在相空间内此类系统的相体积在整个演化过程中保持恒定. 经典力学中非常重要的刘维尔定理表明:哈密顿系统中相体积保持不变,因而说明哈密顿系统是一类保守系统. 刘维尔定理的一个最直接的应用是庞加莱回归定理,该定理表明:保守系统中的轨道从任意点出发,经过充分长时间后将返回或经过该点的任意小的邻域. 上述

定理是保守系统中极少数几个普遍成立的结论之一,具有深刻的意义. 在经典力学中,保守系统分为两大类,即可积系统与不可积系统. 具体地,考虑有 N 个自由度的保守系统,其哈密顿函数为

$$H = H(p_1, p_2, \cdots, p_N, q_1, q_2, \cdots, q_N),$$

其运动方程为

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}.$$

若能找到一系列正则变换 $(p_1, p_2, \cdots, p_N, q_1, q_2, \cdots, q_N) \to (J_1, J_2, \cdots, J_N, \theta_1, \theta_2, \cdots, \theta_N)$,使得新的哈密顿函数只依赖于前一半变量,即

$$H = H'(J_1, J_2, \cdots, J_N),$$

而哈密顿方程为

$$\begin{cases} \dot{\theta}_i = \dfrac{\partial H'}{\partial J_i} = \Omega_i \ (J_1, J_2, \cdots, J_N), \\ \dot{J}_i = -\dfrac{\partial H'}{\partial \theta_i} = 0, \end{cases}$$

则上述哈密顿系统称为可积的. 就可积系统而言,系统在相空间的运动限制在 N 维环面上,而这仅仅是 $2N-1$ 维等势面的一部分,其运动是完全确定的.

余下的另一大类保守系统则是不可积系统. 事实上对 $N \geqslant 2$,任意的保守系统几乎都是不可积的,可积系统是如此稀少,以至于用它们来逼近一个不可积系统也是不可能的. KAM 定理表明:不同于可积系统,不可积系统在相空间中导致不稳定运动的点不必都是测度为零的例外情形,而可能占据有限测度的区域,因而保守系统(不可积系统)亦可存在有不稳定的复杂运动. 保守系统并不局限于经典力学的范畴. 由著名的麦克斯韦方程描述的光在其真空中的传播提供了电磁学领域里保守系统的重要实例.

保守系统(conservative system) 即"守恒系统".

耗散系统(dissipative system) 一类非守恒系统,其背景例如为有摩擦阻尼耗散能量的系统.

KAM 定理(Kolmogorov-Arnold-Moser theorem) 研究近可积系统理论的最重要结果之一. 该定理以三位对此做出突出贡献的数学家的名字命名,他们是苏联数学家柯尔莫哥洛夫(Колмогоров, А. Н.)、阿诺尔德(Арнольд, В. И.)和美国数学家莫泽(Moser, J. K.),从而简称 KAM 定理.

设保守系统为自由度 n,它具有 n 个广义动量 $p_1, p_2, \cdots, p_n$ 和 n 个广义坐标 $q_1, q_2, \cdots, q_n$,以及哈密顿函数

$$H = H(p_1, p_2, \cdots, p_N, q_1, q_2, \cdots, q_N),$$

则系统的运动方程(哈密顿方程)具有正则形式

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \ (i = 1, 2, \cdots, n). \tag{1}$$

对于可积系统,总存在正则变换

$$I_i = I_i(p_1, p_2, \cdots, p_n, q_1, q_2, \cdots, q_n),$$
$$\theta_i = \theta_i(p_1, p_2, \cdots, p_n, q_1, q_2, \cdots, q_n),$$

使得系统的哈密顿函数变成

$$H = \overline{H}(I_1, I_2, \cdots, I_n),$$

而相应的哈密顿方程变为

$$\begin{cases} \dot{I}_i = 0, \\ \dot{\theta}_i = \dfrac{\partial \overline{H}}{\partial I_i} = \Omega_i \ (I_1, I_2, \cdots, I_n). \end{cases} \tag{2}$$

近可积系统是一类不可积系统,它的哈密顿函数可表示为可积哈密顿函数与小扰动项之和,即

$$H = H_0(I_1, I_2, \cdots, I_n) + \varepsilon H_1(I_1, I_2, \cdots, I_n, \theta_1, \theta_2, \cdots, \theta_n), \tag{3}$$

式中,H_0 为可积哈密顿函数,H_1 为不可积哈密顿函数,ε 为一小参数.

KAM 定理描述了近可积系统的定性行为,具体表述为:若 ε 充分小,且未扰动系统 H_0 满足条件

$$\frac{\partial(\Omega_1, \Omega_2, \cdots, \Omega_n)}{\partial(I_1, I_2, \cdots, I_n)} \neq 0,$$

则近可积系统(3)具有勒贝格测度 $\mu(\varepsilon)$ 的运动轨道处于光滑的 n 维环面之上,而这些 n 维环面与可积系统 H_0 的运动轨道所处的光滑环面(其勒贝格测度与 $\mu(I)$)相差很小. 进一步,当 $\varepsilon \to 0$ 时,$\mu(\varepsilon)/\mu(I) \to 1$. 在留存的 n 维环面之间充满着稠密的"无理"轨道. 通俗地说,当 ε 充分小时,系统的绝大多数运动轨道仍处于 n 维环面上,而这些 n 维环面相对于可积系统 H_0 的 n 维环面只有微小的位移而没有定性结构上的改变. 从另一方面看,无论扰动(或 ε)多小,总有一些轨道逃离不变 n 维环面而改变原来的"有理"运动状态,复杂的运动轨道由此产生(尽管当 $\varepsilon \to 0$ 时,这些轨道的测度为 0).

需要指出的是,这些复杂运动的区域不一定是完全孤立的,即两个邻近环面间的轨道有可能相互迁移. 事实上,要保证这些复杂运动的孤立性,必须要求 n 维 KAM 环面成为 $2n-1$ 维等势面的边界,即满足条件

$$n \geqslant 2n - 2, \tag{4}$$

不等式(4)只有在 $n \leqslant 2$ 时成立. 例如,二维环面可以把 $2 \times 2 - 1 = 3$ 维的等能面分割为内、外两部分. 不可积系统中普遍存在着复杂运动,容易设想当系统(3)中 ε 逐渐增大时,KAM 定理所保证存在的 n 维环面将越来越少,随着这些环面的消失,越来越多的复杂运动轨道随之产生,最终形成全局性的复杂运动景象.

中心流形定理(center manifold theorem) 处理非线性系统的一种理论工具. 其基本思路是通过对系统的中心流形的分析和估计来研究系统的定性性质. 下面以有限维自治微分方程系统为例,介绍中

心流形定理的基本内容. 考虑系统

$$\begin{cases}\dot{x} = Ax + f(x,y), \\ \dot{y} = By + g(x,y),\end{cases} \quad (1)$$

式中 $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, A 和 B 均为常值矩阵, 且 A 阵所有特征值均具零实部, B 阵所有的特征值均具负实部. 对一般系统 $\dot{x}=f(x)$, $x\in\mathbb{R}^k$, 其局部不变流形是指集合 $S\subset\mathbb{R}^k$, 其中若 $x(0)=x_0\in S$, 则 $x(t)\in S$, $T>t>0$. 进而若 $T=\infty$, 则集 S 称为不变流形. 进而对系统(1), 若 $y=h(x)$ 为(局部)不变流形且 $h(0)=0, h'(0)=0$, 则称 $y=h(x)$ 为其(局部)中心流形.

关于系统(1)存在中心流形的基本条件有:

1. 假定系统(1)中 f,g 均为 $\mathbb{R}^n$ 或 $\mathbb{R}^m$ 中的 C^r 向量场, $f(0)=g(0)=0$, 则存在 C^{r-1} 的局部中心流形 W^c(即 $y=h(x)$), 它与相应的特征向量空间 E^c(由所有零实部的特征向量构成)相切于 0 点.

利用中心流形, 人们可对系统(1)进行定性分析与计算. 具体地, 将 $y=h(x)$ 代入系统(1)有

$$\dot{x}=Ax+f(x,h(x)). \quad (2)$$

2. 若系统(2)的零解是稳定的(渐近稳定的或不稳定的), 则系统(1)的零解亦是稳定的(渐近稳定的或不稳定的).

中心流形还可通过另一微分方程进行估计. 考虑 C^{r-1} 函数 φ: $\mathbb{R}^n\to\mathbb{R}^m$, 定义算子 M:

$$(M\varphi)(x)=\varphi'(x)(Ax+f(x,\varphi(x))) - B\varphi(x)-g(x,\varphi(x))),$$

则 $\varphi(x)$ 为中心流形 $h(x)$ 的近似估计.

3. 假定 φ 为 $\mathbb{R}^n$ 零点邻域上的 C^{r-1} 的映射, $\varphi(0)=0, \varphi'(0)=0$, 且 $(M\varphi)(x)=O(|x|^q)$, $q>1$, 则 $x\to 0$ 时 $|h(x)-\varphi(x)|=O(|x|^q)$.

中心流形定理还可在下述方面得到推广:

1) 系统(1)线性部分中所有特征值均为非正实部的假定不是必要的.

2) 系统(1)可以不是自治的.

3) 在一定条件下, 平衡点(如零点)可换为不变集.

4) 系统(1)可以是无穷维系统, 如偏微分方程系统.

中心流形定理在突变分析、稳定性分析等方面有广泛的应用.

稳定性(stability) 动态系统的基本性质之一. 指系统行为、现象在起始或外界条件变化下保持稳定不变的性质. 亦即由因果的连续依赖关系所保证的事物性质或现象的相对不变, 即稳定的性质. 常见的稳定性质包括平衡状态的稳定性、运动轨道的稳定性等, 均保证系统行为不会由于偶然及微小的外因变化而破坏. 在数学上, 通常用连续依赖关系的语言来描述, 既可用实分析中的 ε-δ 语言, 也可用点集拓扑的开集结构, 导致各种具体的定义(参见"状态稳定性"、"响应稳定性(输入-输出稳定)").

状态稳定性(state stability) 动态系统状态对初值连续依赖的局域性质. 系统的运动状态 $X(t)$ 随初始条件 $X(t_0)$ 的改变而变. 若对任意给定的 $\varepsilon>0$, 总能找到 $\delta=\delta(\varepsilon)$, 使得当系统初始条件之差

$$|X(t_0)-\widetilde{X}(t_0)|<\delta$$

时, 系统状态之差恒限制在 ε 范围内, 即

$$|X(t) - \widetilde{X}(t)| < \varepsilon, \ \forall t > t_0,$$

则称该系统运动状态 $X(t)$ 是稳定的; 若此外还有

$$\lim_{t\to\infty}|X(t) - \widetilde{X}(t)| = 0,$$

则称系统该状态是渐近稳定的. 特别地, 若 $X(t)$ 是平衡态 $x(t)\equiv x_0$ 时, 则称此平衡态是稳定的.

响应稳定性(response stability) 对系统输入-输出响应特性所定义的稳定性. 设系统输入 $u(t)$ 和输出 $y(t)$ 分别赋予了某种度量 $\|u(t)\|$ 和 $\|y(t)\|$, 特定响应 $y_0(t)=F[u_0(t)]$ 的稳定性可定义为: 任给 $\varepsilon>0$, 必可找到 $\delta>0$, 使得只要 $\|u-u_0\|<\delta$, 就可保证 $\|y-y_0\|<\varepsilon$. 利用系统各部分的输入-输出响应增益特性可以判断整个系统的稳定性, 诸如小增益定理、圆盘定理等.

李亚普诺夫指数(Liapunov exponent) 简称 LE. 刻画动态系统稳定性的重要概念. 在动态系统中判断其解是否稳定的一个方法是: 设某解在初始时刻 t_0 有一扰动 ε_0, 由之引起其后 t 时刻解的扰动为 ε_t, 而 ε_t 和 ε_0 的关系可渐近地用某一常数 LE 来表示为

$$\varepsilon_t \cong \varepsilon_0 \cdot \mathrm{e}^{\mathrm{LE}\cdot t} \ (t>T),$$

若 LE<0, 则 ε_0 是李亚普诺夫意义下稳定的. 由此求出

$$\mathrm{LE} = \lim_{t\to\infty}\frac{1}{t}\ln\frac{\varepsilon_t}{\varepsilon_0}.$$

它表示时间 t 内轨道的平均扩散率(LE>0)或平均收缩率(LE<0), LE 就称为李亚普诺夫特征指数, 简称李亚普诺夫指数. 一般 n 维动力系统可以有 n 个李亚普诺夫特征指数, 各对应于不同的轨道, 并可按大小排列来表征其形态的特征. 例如, 对三维动力系统中的四种不同吸引子, 它们的三个李亚普诺夫指数的符号决定了吸引子的性质. 定常吸引子, 表示三个方向都收缩. 极限环吸引子, 沿环方向既不扩散也不收缩, 其他两个方向都收缩到环. 拟周期二维环面吸引子, 沿环面上两个方向(一个方向绕环面转, 一个方向沿环面前进)不扩散也不收缩, 另一个方向收缩到环面. 混沌吸引子, 既有敏感初条件的伸长性质($\mathrm{LE}_1>0$), 又有折叠性质($\mathrm{LE}_3<0$), 沿流的方向不辐散($\mathrm{LE}_2=0$). 四种吸引子中只有奇怪吸引子中有正的李亚普诺夫指数, 这是混沌吸引子区别于其

他吸引子的独有特征.

同宿点(homoclinic point)　动态系统相点的一种.其特征是沿正、反方向动态演化趋向于同一周期轨吸引子.亦可简称之为同宿轨道上的相点(参见“同宿轨道”).

异宿点(heteroclinic point)　动态系统相点的一种.其特征是沿正、反方向动态演化趋向于不同周期轨吸引子.亦可简称之为异宿轨道上的相点(参见“异宿轨道”).

同宿轨道(homoclinic orbit)　一种特殊相轨道.即动态系统具以下特征的相轨道:它不是奇点或闭轨,但它是沿时间正向演化的 ω 极限集和反向演化的 α 极限集,若它们有相同的奇点或闭轨,则称为同宿轨道;反之,若它们有不同的奇点或闭轨,则称

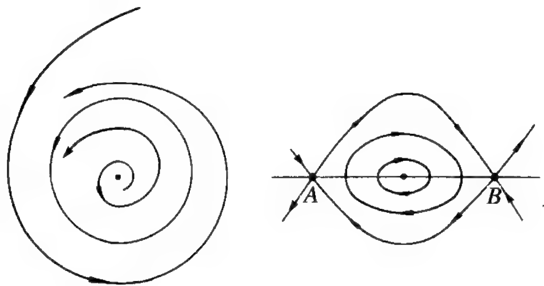


图 1　　　　图 2

为异宿轨道.例如,图 1所示的二维系统有一个稳定极限环和一个不稳定焦点,极限环内的轨道或相点均为异宿的.又如,图 2 所示的无阻尼单摆的大范围相图中,由于 A,B 两点物理上对应于同一个鞍点,连结它们的轨道及其上相点均为同宿的.洛伦兹吸引子对应于一种更加复杂的同宿轨道和同宿点.

异宿轨道(heteroclinic orbit)　见“同宿轨道”.

突变与分岔

突变论(catastrophe theory)　研究控制参数连续变化引起系统状态非连续变化的数学系统理论.突变理论的主要创始人是法国数学家托姆(Thom, R.).突变论的思想可以追溯到法国数学家庞加莱(Poincaré,(J.-)H.)、苏联数学家安德罗诺夫(Андронов, A. A.)的动态系统分岔理论和美国数学家惠特尼(Whitney, H.)的映射奇异性理论.1969年,托姆为解释生物形态发生问题发表论文“形态发生动力学”,提出突变论的基本思想.1969 年,发表“生物学的拓扑模型”,奠定了这一理论的基础.1972年,在专著《结构稳定与形态发生》中系统阐述了突变论.该领域重要代表人物还有英国数学家塞曼(Zeeman, E. C.)、苏联数学家阿诺尔德(Арнольд, В. И.)等.“突变论”一词是由塞曼提出的,他将突变论纳入“系统论”的范畴,从而大大拓宽了突变论可能应用的范围.初等突变论研究有势系统,描述突变现象的基本术语为势函数族、突变流形、奇点集合、分岔点集合等.初等突变论已有成熟的方法和系统的结果,核心是托姆关于突变的完备分类定理.已经证明,当有势系统势函数的控制参数个数 $m\leqslant 4$ 时,至多可有 7 种基本突变类型,如表所示.

余维数	突变芽	标准势函数	突变类型
1	x^3	$ax+x^3$	折叠突变
2	$\pm x^4$	$a_1x+a_2x^2\pm x^4$	尖点突变
3	x^5	$a_1x+a_2x^2+a_3x^3+x^5$	燕尾突变
4	x^2y-y^3	$a_1x+a_2y+a_3y^2+x^2y-y^3$	椭圆型脐点突变
5	x^2y+y^3	$a_1x+a_2y+a_3y^2+x^2y+y^3$	双曲型脐点突变
6	$\pm x^6$	$a_1x+a_2x^2+a_3x^3+a_4x^4\pm x^6$	蝴蝶突变
7	x^2y+y^4	$a_1y+a_2y+a_3x^2+a_4y^2+x^2y+y^4$	抛物型脐点突变

当 $m\leqslant 5$ 时,则有 11 种基本突变.除上述 7 种外,还有印第安人茅舍型突变、第二椭圆型脐点突变、第二双曲型脐点突变、符号型脐点突变等.由于突变是系统演化中常见的现象,代表系统定性性质的改变,在系统研究中有重要意义.突变论已成为建立系统科学基础理论系统学的重要依据之一.初等突变形式可分为冲突型和分支型两类,应用上对力学、物理等对象较为成功,已取得若干定量的结果,用于描述社会问题则基本上还是隐喻式的.高等突变论则研究非有势系统的突变现象.

结构稳定性(structural stability)　系统运动定性特征对其参数微小变化的稳定性.亦即系统运动模式总体形态相对于某结构参数 $d\in D$ 微小变化保持不变的稳定性问题.如果一个(连续流或离散的)微分动力系统在“小扰动”下其轨道相图的拓扑结构不变,则称这个微分动力系统是结构稳定的.结构稳定性是微分动力系统的中心论题之一.与结构稳定相对的概念,即结构不稳定性就是分岔的概念.对于离散的微分动力系统,结构稳定性可定义为:设 $\mathrm{Diff}^r(M)$为 C^r 微分流形 M 上所有 C^r 微分同胚构成的集合,并赋予映射空间意义下的 C^r 拓扑,离散动力系统 $f\in\mathrm{Diff}^r(M)$称为是 C^r 结构稳定的,如果存在 f 在 C^r 拓扑中的邻域 $U\subset\mathrm{Diff}^r(M)$,使得任意的动力系统 $g\in U$ 都与 f 拓扑共轭,即对任何 $g\in U$ 都存在 M 的自同胚 h,使得 $h\circ f=g\circ h$.上述定义中应用了拓扑共轭这一概念来描述动力系统有相同拓扑结构这一说法.拓扑共轭是一等价关系.在这一

等价关系之下，$\mathrm{Diff}^r(M)$中的全体动力系统被分成拓扑等价类. 如果在上述定义中减弱拓扑共轭的关系，使之成为在C^r微分流形M上动力系统f的某一不变集Λ上或某一点邻域的局部拓扑共轭，则分别得到动力系统f在不变集Λ上或该点的局部结构稳定的定义. 最常考虑也是极有意义的情形是动力系统在其非游荡集(或称Ω集)上的结构稳定性. 但是美国数学家斯梅尔(Smale，S.)证明，在一个四维流形M^4上，存在一个向量场(M^4, X)，它无法用结构稳定场逼近，这就表明前面的想法只适用于一定范围.

分岔理论(bifurcation theory) 研究动力学系统分岔现象的数学理论. 分岔指非线性系统的控制参数变化于某些临界点时引起的系统定性性质的改变，包括稳定性的变换、定态个数的增减、类型的改变、相空间中定态分布的改变等. 控制空间中发生分岔的参数点称为分岔点，分岔点的集合构成的曲线或曲面，称为分岔曲线或分岔曲面. 分岔是非线性动力学系统的通有现象，是系统通过演化产生多样性、复杂性的重要机制. 分岔理论的主要内容是：分岔的定义，分岔发生的条件，分岔类型，求分岔解或其解析表达式的方法，分岔解的稳定性分析，阐明各类分岔所代表的系统行为等.

隐函数定理(implicit function theorem) 由隐函数形式求解因变量对自变量显式依赖关系的方法和可解性定理. 设变量x, μ满足方程$f(x,\mu)=0$，其中μ是控制参数，μ和x在方程中存在隐函数关系. 若在某点(x_0,μ_0)附近有

$$\frac{\partial f}{\partial \mu} \neq 0,$$

则由

$$\mathrm{d}f = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial \mu}\mathrm{d}\mu = 0$$

可推出

$$\frac{\mathrm{d}\mu}{\mathrm{d}x} = -\left(\frac{\partial f}{\partial x}\right)\Big/\left(\frac{\partial f}{\partial \mu}\right),$$

即在(x_0,μ_0)附近存在惟一解$\mu(x)$. 隐函数定理特别指出：若$\frac{\partial f}{\partial \mu}=0$，则隐函数方程$f(x,\mu)=0$，解的惟一性可能被破坏，从而在点$(x_0,\mu_0)$将产生分岔.

同样对于多维非线性动力系统

$$\dot{x} = f(x,\mu),$$

其定常解满足隐函数方程

$$f(x_0,\mu_0) = 0.$$

若

$$\left.\frac{\partial f}{\partial x}\right|_{x_0,\mu_0} = 0,$$

即雅可比矩阵的特征值实部为零，则(x_0,μ_0)是一个分岔点. 又如对非线性迭代映射

$$x_{n+1} = f(x_n,\mu),$$

令$g(x,\mu)=f(x,\mu)-x$，则映射不动点满足

$$g(x,\mu)=f(x,\mu)-x=0.$$

又若

$$\left.\frac{\partial g}{\partial x}\right|_{x_0,\mu_0} = \left.\frac{\partial f}{\partial x}\right|_{x_0,\mu_0} - 1 = 0,$$

则上式表明雅可比特征值的模(乘子)在单位圆上，从而点(x_0,μ_0)可能是分岔点.

余维数(codimension) 突变论中描述其势函数性质的一个整数. 设$f: \mathrm{R}^n \to \mathrm{R}$为$n$维空间中的光滑函数. 记$\varepsilon(n)$为$\mathrm{R}^n \to \mathrm{R}$的光滑函数的全体，$\varepsilon(n, m)$为$\mathrm{R}^n \to \mathrm{R}^m$的$m$维光滑函数的全体，$m(n)$为$\varepsilon(n)$中满足$f(0,\cdots,0)=0$的函数子集. 以

$$\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right)$$

为生成元，以$\varepsilon(n)$为系数集的代数环记为Q. 在$m(n)$中定义等价类如下：$f \sim g$，若$f \sim g \in Q$，则以这些等价类为基元构成的向量空间的维数就定义为f的余维数.

在突变理论中，余维数就等于势函数的万能扩展所需引入的控制参数的个数.

扩展(unfolding) 突变论术语. 指通过增加扩展参量将多元函数嵌入某函数族的过程. 多元函数$f(x_1,x_2,\cdots,x_n)$的一个r次扩展是增加了r个扩展参量$t_1,t_2,\cdots,t_r$的多元函数$F(x_1,x_2,\cdots,x_n;t_1,t_2,\cdots,t_r)$，且满足

$$F(x_1,x_2,\cdots,x_n;0,\cdots,0) = f(x_1,x_2,\cdots,x_n).$$

通过这样把单个函数嵌入某个函数族，有可能把一个结构不稳定的函数嵌入一个几乎处处结构稳定的函数族中. 这种能使函数族稳定化的扩展称为完全扩展，而其中包含参量最少的扩展称为万能扩展. 万能扩展的参量个数就是余维数. 一个简单例子为：对$f(x)=x^4$，$F_1=x^4+ax^2$是个扩展，但不是完全扩展；$F_2=x^4+dx^3+ax^2+bx$和$F_3=x^4+ax^2+bx$都是完全扩展，且F_3还是万能扩展，其余维数是2. 扩展方法被用于突变理论中引入适当的控制参数，并得到了基本突变的完备分类.

***k*阶截断**(k-jet) 突变理论的运算步骤. 指多元函数泰勒展开截取若干低阶项以简化处理的方法. 多元函数$f(x_1,x_2,\cdots,x_n)$在原点的k阶截断是f在原点的泰勒展开保留k阶以下的多项式部分，记为$J^k f$. 截去的部分称为f的尾部，记为$\mathrm{Tayl}\,f$. 从而有

$$f = J^k f + \mathrm{Tayl} f.$$

在突变理论中，利用k阶截断把任意光滑势函数化为多项式函数族进行研究，得到了基本突变的完备分类.

突变芽(germ)　突变理论术语.指势函数分解之后决定突变特征的部分.在突变理论中,按剖分引理可以把势函数在退化临界点处分解为莫尔斯(非退化)部分和一个最小的非莫尔斯部分,即 $V=M+G(r)$.后者 $G(r)$ 即为突变芽,其实质性变量个数为余维数 r.法国数学家托姆(Thom,R.C.)的基本突变分类问题实质上就是对突变芽进行的分类.

塞曼突变机(Zeeman catastrophe machine)　一种直观的突变机构模型.由英国数学家塞曼(Zeeman,E.C.)于1972年创造(如图).轻质圆盘由两根弹性索 AB 和 BC 牵引,圆心 O 和 A 点固定在平面上,而 B 点固定在圆盘上.当 C 点移动时,B 点将有不同的平衡位置.其特殊之处在于:当 C 点沿不同方向到达某些特定位置时,B 点会从左侧或右侧的连续变化的平衡位置突跳到另一侧去.这一模型为托姆突变理论提供了一个直观的例子,对普及和理解突变理论起了较大作用.详细的数学分析可以证明:塞曼突变机的势函数为

$$V=\frac{1}{4}x^4+\frac{1}{2}ax^2+bx,$$

具有尖点型突变.

余秩(corank)　多元函数黑塞矩阵不满秩程度的指标.光滑函数 $f(x_1,x_2,\cdots,x_n)$ 的黑塞矩阵 Hf 的零特征根的个数 r,称为函数 f 或矩阵 Hf 的余秩,$n-r$ 为秩.$r=0$ 表示黑塞矩阵满秩,代表非退化定态点,不存在退化现象.$r\neq0$ 代表退化定态点,黑塞矩阵非满秩,表示在状态空间的 n 个坐标方向中有 r 个是退化方向.由 r 个退化方向张成的空间,称为 f 的零空间,余秩 r 是零空间的维数.

莫尔斯引理(Morse lemma)　关于势函数在非退化定态点附近定性性质的重要命题.设 $f:\mathrm{R}^n\to\mathrm{R}$ 为一光滑函数,u 是它的一个非退化定态点,则在 u 的某个邻域 U 内存在一个局域坐标变换 $Y=(y_1,y_2,\cdots,y_n)$,满足 $y_i(u)=0(i=1,2,\cdots,n)$,使得对于一切 $u\in U$,函数 f 在此局域坐标中变为二次型

$$f(Y)=-y_1^2-y_2^2-\cdots-y_k^2+y_{k+1}^2+\cdots+y_n^2,$$

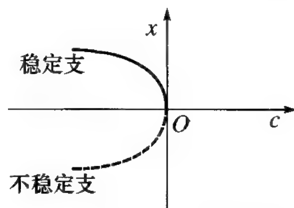
称为莫尔斯标准型,或莫尔斯 k 级鞍.式中负项个数 k 是一个重要参数,称为非退化定态点的指数.k 为拓扑不变量,不因坐标变换而改变.莫尔斯引理保证每个非退化定态点通过一个光滑可逆坐标变换为一个 k 级莫尔斯鞍点.k 满足条件 $0\leqslant k\leqslant n$,$k=n$ 时势函数在0点取极大值,$k=0$ 时势函数在0点取极小值,都不是真鞍点.$0<k<n$ 时0点才是鞍点.对于一元势函数 $V(X)$,引理保证它在莫尔斯点附近与抛物线函数 $V=\pm y^2$ 有相同的定性性质.

剖分引理(splitting lemma)　关于势函数在退化定态点附近定性性质的重要命题.在退化定态点上,势函数的一阶导数和黑塞矩阵的行列式同时为零,莫尔斯引理条件不成立,势函数不能表为二次型.这时,若0点为光滑函数 $f:\mathrm{R}^n\to\mathrm{R}$ 的退化定态点,f 在0点的黑塞矩阵的余秩为 $n-r$,则剖分引理保证 f 在0点附近可表示为

$$f\triangleq f_{NM}(y_1,y_2,\cdots,y_r)\pm y_{r+1}^2\pm\cdots\pm y_n^2,$$

其中 $f_{NM}(y_1,y_2,\cdots,y_r)$ 为光滑函数,称为函数 f 的非莫尔斯部分.n 元势函数在退化定态点附近的奇异行为,一般只与部分状态变量有关.剖分引理表明:势函数在退化定态点上可分为两部分,同奇异性无关的是莫尔斯部分,等价于 $\pm y_{r+1}^2\pm\cdots\pm y_n^2$;同奇异性有关的是非莫尔斯部分.突变理论关注的正是非莫尔斯部分,其中包含的 r 个变量称为实质性变量,是产生突变的因素;其余 $n-r$ 个变量为非实质性的,可以通过坐标变换约去,从而把 n 维问题降低为 r 维问题.这使剖分引理成为研究突变现象的有力工具.

折叠突变(fold catastrophe)　一种最简单的初等突变.折叠突变的势函数为 $V(x,c)=x^3+cx$,状态空间为 x 轴,控制空间为 c 轴,构成乘积空间 xc 平面,其余维数和余秩均为1.突变流形 M 为方程 $3x^2+c=0$ 所确定的 xc 平面上的一条抛物线,如图所示.突变流形上满足方程 $6x=0$ 的点构成奇点集 S,即点(0,0).势函数只有一个孤立奇点是折叠突变的一个特点.奇点集 S 在控制空间(即直线 $x=0$)上的投影,称为分岔点集.折叠突变只有一个分岔点,即 $c=0$.分岔点将控制空间分为两部分,$c>0$ 时势函数无定态点,与线性函数局域同类型;$c<0$ 时势函数有两个定态点,一个稳定,一个不稳定,与抛物线函数同类型.



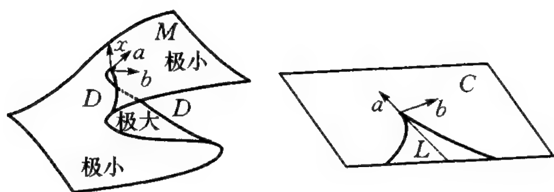
尖点突变(cusp catastrophe)　一种初等突变.它的势函数为

$$V(x;a,b)=x^4+ax^2+bx,$$

状态空间即 x 轴,是一维的,控制空间为二维平面 (a,b).突变流形的方程为

$$\frac{\mathrm{d}V}{\mathrm{d}x}=4x^3+2ax+b=0,\qquad(1)$$

它是三维空间 (x,a,b) 中的一张三叶折叠曲面,如



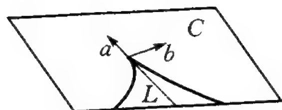


图 1　　　　图 2

图 1、图 2 所示. M 的上、下叶为极小点，势函数稳定；中叶为极大点，势函数不稳定. 对方程(1)两边求导数，得

$$\frac{d^2V}{dx^2} = 12x^2 + 2a = 0. \tag{2}$$

解联立方程组(1)和(2)，得到奇点集 S 应满足的方程

$$8a^3 + 27b^2 = 0. \tag{3}$$

S 代表突变流形 M 的上、中、下三叶之间的两条分界线. 把 S 投影到控制平面 ab 上，得到分岔点集，即控制平面 ab 上由(0,0)引出的两支曲线. 点(0,0)是分岔曲线的一个不光滑的尖点，这是尖点型突变名称的由来；也是势函数奇点的代数类型决定奇点稳定分岔的一种. 尖点突变模型有广泛的应用.

突变特征(catastrophe character)　突变论术语. 指初等突变现象共有的基本特征，也是鉴别系统是否存在突变的依据. 诸如：

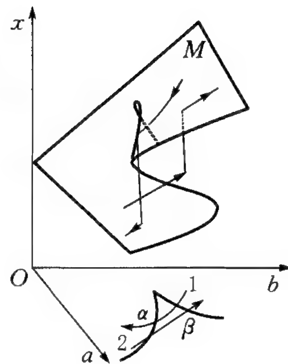
1. 多模态. 存在突变现象的动力学系统一般具有两个或多个可分辨的稳定定态，即在控制空间的一定范围内，对应于同一点(同一组控制参数)的势函数有多于一个的极小点. 如尖点突变中在三叶折叠区具有双模态，控制平面的一点对应两个稳定定态，一个在上叶，一个在下叶. 复杂系统可能有多于两个的模态，只有折叠突变为单模态.

2. 不可达性. 突变流形上的极大点，如折叠突变的 M 曲线下半支上的点，尖点突变的 M 曲面中叶上的点，由于它们是不稳定的，代表系统实际不可能达到的定态.

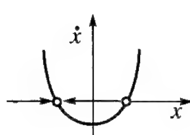
3. 突跳. 在分岔点上，控制参数的微小变化将引起系统从一种稳定定态突然跳跃到另一个稳定定态. 如在尖点突变中从上(下)叶突跳到下(上)叶.

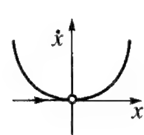
4. 滞后. 反映突变过程具有不可逆性或方向性的现象. 以尖点突变为例，如图所示，控制参数在 xa 平面上变化到达分岔曲线时是否引起突变，与变化的方向有关：当沿路径 1 从右向左变化，在到达分岔曲线的右支时不发生突跳，要到达曲线左支的 α 点才会突跳；反之，当沿路径 2 从左向右变化到达分岔曲线左支时不会突跳，要到达右支的 β 点才会突跳.

5. 发散. 在退化定态点附近控制参数的小变化可能导致系统定性性质的改变，称为发散. 如在尖点突变中，控制参数从分岔曲线内(外)侧一点变到外(内)侧时，可能引起系统状态在上、下叶之间的交换.



鞍结分岔(saddle-node bifurcation)　一种分岔类型. 指控制参数变化过程中系统因形成鞍结点而出现的分岔. 以一阶系统 $\dot{x}=a+x^2$ 为例，控制参数 a 可取任何实数. $a>0$ 时不动点方程 $a+x^2=0$ 没有实数解，表示系统没有定态(图 3). $a<0$ 时方程有两个实





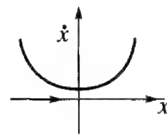


图 1　　　图 2　　　图 3

数解 $x_{1,2}=\pm\sqrt{-a}$，左侧为稳定定态点，右侧则为不稳定定态点(图 1). 图 2 所示分岔点 $a=0$ 处不动点 $x^*=0$ 是半稳定的，称为鞍结点. 从而在 $x\dot{x}$ 平面中用几何方法可方便地找到不动点并判别稳定性.

总之当控制参数 a 由正值逐步减小而跨越鞍结点时，将出现定态点的创生，反方向变化则出现定态点的消失，都表示系统的分岔行为. 这种分岔的特点亦可用图 4 表示.

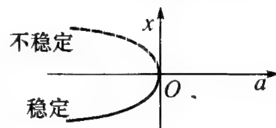
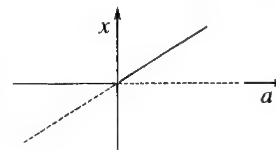
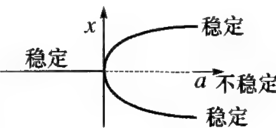


图 4

跨临界分岔(transcritical bifurcation)　一种分岔类型. 指控制参数跨越临界点时出现的方程不同解的稳定性交换. 以方程 $\dot{x}=ax-x^2$ 为例. 对于任何参数 a，都有两个不动点 $x_1=0$ 和 $x_2=a$，系统不存在定态点的创生或消失，分岔表现为两个解的稳定性交换. $a<0$ 时，$x_1=0$ 稳定，$x_2=a$ 不稳定；$a>0$ 时，$x_1=0$ 不稳定，$x_2=a$ 稳定，即稳定性发生交换. 故 $a=0$ 为分岔点. 其分岔特性如上图所示.



叉型分岔(pitchfork bifurcation)　一种分岔类型. 由其几何形状类似叉子而得名. 如图所示. 以方程 $\dot{x}=ax-x^3$ 为例. 控制参数 $a<0$ 时只有一个实数解 $x=0$，代表系统的稳定定态. $a>0$ 时有 3 个不



动点，$x_1=0$ 为不稳定定态，$x_2=\sqrt{a}$ 和 $x_3=-\sqrt{a}$ 为稳定定态. $a=0$ 为分岔点，控制参数由负值增大到跨越这一点时，系统出现定态点的创生，同时发生稳定性的转变.

霍普夫分岔(Hopf bifurcation) 一种分岔类型. 指由参数变化导致稳定焦点失稳而产生周期振荡的分岔，依提出该概念的德国数学家霍普夫(Hopf,E.)的名字而命名. 分岔前的定态为稳定焦点，在分岔点上失稳，并出现一个稳定的极限环，如图所示. 对动力学方程的线性稳定性分析发现，此时其线性近似有两个复共轭特征根

$$\lambda_1=\lambda+\mathrm{i}\omega,\quad \lambda_2=\lambda-\mathrm{i}\omega$$

(其中 λ,ω 为不等于零的实数)穿过虚轴. 稳定焦点通过分岔而失稳，生出一个稳定的极限环，表示系统由平衡态转变为周期振荡.

倍周期分岔(period doubling bifurcation) 一种分岔类型. 指由[0,1]区间迭代时随参数 λ 变化而导致周期倍化而形成的分岔. 动力系统中一种常见且目前研究较多的分岔形式. 以典型的逻辑斯谛迭代过程为例. 其方程为

$$x_{n+1}=f_\lambda(x_n)=\lambda x_n(1-x_n)\quad (\lambda\in[0,4],x_n\in[0,1]),\tag{1}$$

当 $\lambda\leqslant 1$ 时，由于 $x_{n+1}\leqslant\lambda x_n<x_n(x_n\neq 0)$，因此，不论初始 x_1 为何值，当 $n\to\infty$ 时，x_n 必将趋于 0，从而 $x=0$ 是其惟一的稳定不动点.

当 $1<\lambda<3$ 时，其不动点 x^* 满足方程

$$x^*=\lambda x^*(1-x^*),$$

由此可解出 2 个不动点

$$x^*=0 \text{ 和 } x^*=1-\frac{1}{\lambda}.\tag{2}$$

根据不动点的稳定条件

$$|f'_\lambda(x)|<1 \text{ 即 } |\lambda-2\lambda x|<1,\tag{3}$$

可以判断 $x^*=0$ 为不稳定不动点，而 $x^*=1-1/\lambda$ 为稳定不动点.

当 $\lambda>3$ 时，不动点 $x^*=1-1/\lambda$ 亦变为不稳定的. 进一步分析表明，当 $3<\lambda<1+\sqrt{6}=3.449\cdots$ 时，系统的稳定定态为 2 点周期运动. 例如，$\lambda=3.2$ 时，系统解经过一段过渡状态之后呈现出两点交替

$$\begin{matrix}0.5130\\0.7935\end{matrix}$$

的振荡状态. 随着 λ 的进一步增大，系统解将历经 4 点周期状态、8 点周期状态、16 点周期状态等. 将参数值 λ 作横轴，以竖轴表示系统稳态解，则可得到系统(1)的完整的分岔图象(参见“逻辑斯谛映射”图 2). 图中，λ_1 表示单点周期变为 2 点周期的分岔值，λ_2 表示 2 点周期变为 4 点周期的分岔值. 一般地，λ_n 表示 2^{n-1}点周期变为 2^n 点周期的分岔值. 由于每次分岔，系统稳态解的周期增大一倍，故这种分岔过程称为倍周期分岔.

逻辑斯谛映射属于一类非常重要的映射——单峰映射，它的基本特点是 f 和 f' 均为连续的，由某区间 I(本例为[0,1])到自身上的映射，且 f 在区间 I 上有且仅有一个极大值.

对于这类映射，系统(1)中发现的倍周期分岔结构具有普遍意义. 进一步的研究表明，对于这类映射，许多有关倍周期分岔过程的定性定量性质与函数 f 的具体形式无关，表现出一种普适性. 主要结果包括：

1. 单峰映射最多只有一个稳定周期. 对于确定的参数值 λ，从不同的初值 x_0 出发的迭代轨道 $\{f_\lambda^i(x_0)\}$ 可能有许多不同的周期解，但大多数是不稳定的. 因此，绝大多数运动轨道或者趋于同一个稳定的多点周期解，或者进入永不循环的无穷序列(即非周期解).

2. 沙可夫斯基定理. 如果单峰映射 $f_\lambda(x)$ 具有 p 周期点，p 居于下列整数序列(沙可夫斯基序列)之中：

$$\begin{aligned}&3,5,7,9,\cdots\\&3\times2,\ 5\times2,\ 7\times2,\ 9\times2,\ \cdots\\&3\times2^2,\ 5\times2^2,\ 7\times2^2,\ \cdots\\&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\&\cdots,2^m,\ 2^{m-1},\ \cdots,\ 16,\ 8,\ 4,\ 2,1,\end{aligned}$$

则对于序列中位于 p 后的任何整数 q，f 必具有 q 周期点.

3. 分岔收敛速率(费根鲍姆数). 前面用 $\lambda_1,\lambda_2,\cdots,\lambda_n,\cdots$ 顺序表示倍周期分岔序列中的各个分岔值. 随着 n 的增加，两个相邻的分岔点的距离近似地按几何级数形式递减，并最终收敛于 0，而分岔参数值则趋于 λ_∞.

费根鲍姆(Feigenbaum,M.J.)通过大量数值计算首次发现这一规律，并了解到这一规律具有一定的普遍性. 精确地，对于具有二次极大值的单峰映射，各 λ_i 之间有极限关系

$$\lim_{n\to\infty}\delta_n=\lim_{n\to\infty}\frac{\lambda_n-\lambda_{n-1}}{\lambda_{n+1}-\lambda_n}\approx 4.66920\cdots,$$

该极限值称为费根鲍姆常数，对一大类系统具有普适性. 除了单峰映射外，人们在许多其他系统(如高维系统)中也发现了倍周期分岔序列，如受迫布鲁塞尔器、机械碰撞过程、偏微分方程等系统. 显然，倍周

期分岔是一极为普遍的分岔现象.

上述的所有结果或者出自理论分析或者源于数值计算,但是有必要注意的是物理学家们也在实际系统的实验中多次发现并仔细研究了倍周期分岔现象,这方面的工作无疑有助于缩小数学理论与现实世界之间的鸿沟.

状态变量(state variable) 完整描述系统运动的一组变量.它应能确定系统未来的演化行为.例如,理想气体的状态变量为温度 T、压力 P 和体积 V,一维质点运动的状态变量为它的位置 x 和速度 $v=\dot{x}$.对应于系统是连续或跳跃式的演化,其状态变量亦可以是连续的或者是离散的.

控制参数(control parameter) 影响系统状态变量演化特征的一些物理参数.与状态变量比,这些参数变化缓慢,从而它们在系统数学模型中常可作为常数处理,以反映系统对环境的依存制约关系,或可由人从系统外部进行调整,故称为控制参数.以控制参数为坐标轴张成的空间,称为控制空间.控制参数的改变一般只导致系统状态的渐变,但在某些关键点上将引起系统定性性质的突变.

控制参数对系统形态的演化起着决定性的作用.如流体力学的纳维-斯托克斯方程,它描述粘性流体运动速度场 $\boldsymbol{v}(u,v,w)$.控制速度场演变的参数是一维控制空间的雷诺数 Re,当 Re 足够小时流体运动是层流,当 Re 超过临界雷诺数后,运动就变成湍流.

控制空间(control space) 见"控制参数".

庞加莱映射(Poincaré map) 由相空间中轨线运动定义的一种映射.当轨线反复穿越同一截面时,反映后继点对先行点依赖关系的映射.一个连续非线性动力系统的求解是非常困难的,法国数学家庞加莱(Poincaré,(J.-)H.)给出了相图分析法.在相图中虽然不能定量地知道物理量随时间的变化,但可以定性地得到轨线的形态类型及其拓扑结构,从而了解动力系统运动的全局图象.为了更清楚了解高维相空间运动的形态,庞加莱在连续运动的轨线上用一个截面(称庞加莱截面)将其横截,轨线在截面上穿过的情况就可以简捷地判断运动的形态.以 S 记庞加莱截面,$x_n(n=0,1,2,\cdots)$为轨线前一次穿过 S 的点,x_{n+1}为轨线后一次穿过 S 的点,x_{n+1}可看成 x_n 的一种映射:

$$x_{n+1}=f(x_n)\quad(n=0,1,2,\cdots),$$

这个映射就称为庞加莱映射.这样就可从概念上把一个连续的运动化简为离散映射来研究.

同时,该映射的不动点则反映相空间的周期运动.如果运动是二倍周期的,则在庞加莱截面有两个不动点;如果运动是四倍周期的,则有四个不动点等.

哈密顿系统(Hamilton system) 由经典牛顿力学导出的一种规范形式.该系统由美国数学家哈密顿(Hamilton,W.R.)于 19 世纪提出,并与其对偶的变分原理一起被广泛用于物理学的其他领域.其基本的正则方程式为

$$\dot{p}=-\frac{\partial H}{\partial q},\quad \dot{q}=\frac{\partial H}{\partial p}.$$

哈密顿系统事实上可用来描述任何可忽略耗散效应的物理过理,诸如半导体、超导、等离子体、天体力学、材料、结构生物学等领域,包括运动物体控制、光学、等离子体约束、流体和弹性体力学、量子力学、量子场论、广义相对论、孤子与非线性波等科学和技术问题.总之,哈密顿系统将不同学科的规律用统一的数学形式来描述,引起了广泛的关注.

哈密顿体系研究的重要途径是辛(Symplectic)几何,它的计算方法为辛差分格式,它可保持长期的稳定性.中国数学家冯康等人在这方面做出了重要的贡献.

标度理论(scaling theory) 研究相变点附近奇异行为的一种标度变换的理论.例如,在随机游动问题中时间 t 和距离 ξ 之间的关系为 $t\propto\xi^2$,从而 ξ 乘以 $\lambda^{1/2}$的标度变化对应于 t 放大 λ 倍,这种关系称为标度律.一般地,可有 $t(\lambda\xi)=\lambda^{\alpha}t(\xi)$,其中 α 称为标度指数,可用以控制该问题的奇异性.

对非平衡系统,若在某参数值系统具稳定解,则可在定态解附近进行线性化.但在临界点处若参数改变 ε,某些失稳模态对 ε 展开时,可以改为以 ε^2 为时间标度,而 ε 为空间标度.这时仍可能把非线性方程之解展开为临界点处线性方程解的叠加,从而给出非线性方程的一种求解途径.

耗散结构理论与协同学

耗散结构理论(dissipative structure theory) 研究远离平衡态的开放系统从无序到有序演化规律的一种理论.其基本要点为:一个远离平衡态的开放系统,通过不断与环境交换物质与能量,当系统的某些参量变化达到一定临界值时,由于涨落引起系统突变(非平衡相变),使系统由原来的无序状态转变为一种在时间、空间或功能上的有序结构,即耗散结构.之所以称为耗散,是由于这种结构须不断与环境交换物质和能量才能产生和维持.这是一种无需外界提供结构信息而自行产生具有某种特征结构的自组织过程.因此,耗散结构理论也是系统自组织理论的一种.它是由比利时学者普利高津(Prigogine,I.)于 1969 年提出的.耗散结构理论首先在贝纳德对流、贝洛索夫-扎波金斯基反应、激光形成等物理、化学实验中得到证实,之后又被推广到对生物、生态乃

至社会现象的理解，用远离平衡态的开放系统通过涨落作用形成高度有序的耗散结构来解释各种复杂的自组织进化过程. 耗散结构理论对上述各类系统的自组织过程进行了数学描述和详细的物理分析. 所用的理论方法涉及动态系统的稳定性和极限行为、分岔和突变理论、随机过程以及非平衡统计物理的许多成果. 由这些实证研究可以概括出耗散结构的形成须有以下几方面条件：

1. 系统必须是开放的，由于开放系统与外界有物质及能量的交换而带来负熵流，可以导致总熵的减少和系统的有序化，而孤立系统的熵只会不断增加而趋于无序的平衡态.

2. 系统必须是非线性的，其内部各要素可产生协同和相干效应，以形成多种多样的吸引子结构.

3. 系统必须远离平衡态，才能脱离原来的吸引子而形成新的结构.

4. 在上述条件下，通过涨落触发突变并被放大为宏观的结构变化而达到新的有序.

阿尼斯定理(Hanus theorem)　刻画化学反应数学模型中奇点性质的定理. 该定理断言：一个包含两个中间产物的化学反应序列中，如果各个反应只有单分子或双分子参加，则系统不可能有包围不稳定结点或焦点的极限环. 该定理首先由阿尼斯(Hanus)于 1972 年提出. 他对空间均匀、无扩散情况给出了证明，后人将其推广至空间不均匀、存在扩散的情况.

阿尼斯定理告诉人们，为使一个反应系统出现极限环表示的有序结构，其中必须有一个三分子参加的反应. 正是基于此分析，比利时布鲁塞尔学派提出著名的三分子模型(参见“布鲁塞尔模型”).

贝纳德对流(Benard convection)　一种流体自组织现象. 由法国人贝纳德(Benard,C.)于 1990 年发现. 当由底部加热水平金属板上的流体薄层时，开始只有微观的热传导而宏观上保持静止. 但当温度梯度超过某临界值时，流体会突然出现宏观可见的对流图案结构. 当上表面为自由时，从上向下可见其形状为六角形格子；而上表面亦有平板约束时，从侧面观察则可见对流呈两两相背方向的旋转卷筒状. 这种现象称为贝纳德对流. 后人进一步研究发现，其失稳条件为瑞利数

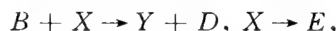
$$Ra = \frac{g\beta}{kv}d^2\Delta T$$

达到某临界值，其中 g 为重力加速度，β 为流体体积膨胀系数，k 为热导率，v 为运动粘性系数，d 为流体层厚度，ΔT 为流体上下面温差值. 贝纳德对流是非平衡系统自组织或耗散结构现象的早期例子(参见“耗散结构理论”).

布鲁塞尔模型(Brusselator model)　亦称三分子模型. 由布鲁塞尔学派提出模拟自组织现象的一种数学模型. 其三分子反应模式为



$$A \to X,\ 2X + Y \to 3X,$$



$$B + X \to Y + D,\ X \to E,$$

这里 A,B 为初始反应物，D,E 为最后产物，X,Y 为中间组分. 设反应过程中 A,B,D,E 保持不变，而 X,Y 的浓度随时间变化，可以得到系统的动力学方程为

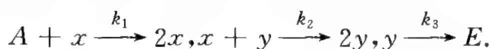
$$\frac{dX}{dt} = A + X^2Y - BX - X,$$

$$\frac{dY}{dt} = BX - X^2Y.$$

它的定态解为 $X_s = A, Y_s = B/A$. 当控制参数 $B > 1 + A^2$ 时产生失稳，出现一个稳定极限环. 这一模型可用来解释化学钟、生物钟等现象，有重要的理论意义.

三分子模型(three molecule model)　即“布鲁塞尔模型”.

洛特卡-沃尔泰拉方程(Lotka-Volterra equation)　一种描述生态演化过程的数学模型. 设两种鱼 X,Y 数量分别为 x,y. 鱼 X 吃微生物 A，而鱼 Y 吃鱼 X. 其数量演化可用下列反应式表示：



$$A + x \xrightarrow{k_1} 2x,\ x + y \xrightarrow{k_2} 2y,\ y \xrightarrow{k_3} E.$$

且其动力学方程为

$$\frac{dx}{dt} = k_1Ax - k_2xy,$$

$$\frac{dy}{dt} = k_2Axy - k_3y.$$

洛特卡-沃尔泰拉方程有不稳定的平衡解 $x_0 = 0, y_0 = 0$；另一平衡解

$$x_s = \frac{k_3}{k_2},\quad y_s = \frac{k_1A}{k_2}$$

则是中心型的，围绕它有一族周期轨道. 任何小的扰动可使它由一个周期轨道跳到另一个上去. 这描述了自然界生态系统的周期变化，但不形成耗散结构.

对称破缺(symmetry breaking)　一种系统演化机制. 系统的对称指其在某种变换(平移、旋转、折叠等)下的不变性，亦即不可区分性，是无序的表现. 反之，对称破缺则意味着对称性的减少，亦即有序结构的形成. 此外，演化过程的不可逆性，也是时间反演的对称破缺.

平衡态(equilibrium state)　统计物理学中假定的一种理想状态. 此时系统的所有状态变量都不随时间变化，系统不同位置上的状态也完全均匀. 与外界环境无任何形式、物质、能量交流的孤立系统在无限长的时间以后，系统所达到的状态可以认为是平衡态.

非平衡态(non-equilibrium state)　系统非均

匀或随时间变化的状态. 当系统与外界存在物质或能量交流时,其状态为非平衡态.

非平衡态又可分为两类:

1. 系统与外界有交流,但其状态变量并不改变,这时称系统处在非平衡定态,这时系统内部存在输运现象.

2. 从系统某些位置吸收的物质或能量,经过输运又从系统的另一位置流入环境. 例如,生物体不断吃进食物,吸进氧气,变成热量,又经运动而排出体外,从而维持于一个非平衡定态.

负熵(negentropy or negative entropy) 一种取负值的熵. 减少系统熵的情况即可视之为获得了"负熵",这往往又可理解为获取了信息.

物理学家薛定谔(Schrödinger,E.)在《生命是什么》一书中利用负熵概念来解释生命现象. 尽管按热力学第二定律,人体内自发过程导致熵增加,但实际上人体仍能保持在一个低熵的状态. 由此可以认为,存在一个人体与外界进行物质、能量交换的过程,使其熵减少,并最终使人体维持在一个低熵状态. 薛定谔形象地将之比喻为"生命就是负熵". 这是开放系统自组织过程研究在生物方面应用的一个重要的问题.

孤立系统的熵增加原理(principle of entropy increasing in isolated system) 热力学第二定律的一种表述形式. 一个与外界无任何物质、能量交流的孤立系统,其演化过程中,熵总是在增加,直至达到熵最大的平衡态. 数学上这一过程可表示为 $dS \geqslant 0$.

相变(phase transition) 系统结构或形态的定性变化. 指随着温度、压强等物理参数达到一定临界值时,物质的相结构发生的突变. 常见的相变结构包括诸如物质的气态、液态和固态的相互转变,铁磁性物质达到居里点时磁性的出现或消失等.

昂萨格倒易关系(Onsager reciprocal relation) 近平衡热力学系统流与力之间比例系数的一种关系. 当系统离开平衡态时,维持系统处在非平衡状态的外界条件,如温度差、浓度差、电位差等广义力可记为 X_i,它们引起系统内部产生相应的流,如热量流、物质流、电子流等,可称之为广义流,并记为 J_i. 在离平衡态不远时,流与力成正比. 昂萨格倒易关系概括了 19 世纪以来的一大批物理关系,如欧姆定律(电流与电压成正比)、傅里叶热传导定律(热流正比于温度差)、菲克扩散定律(粒子流正比于浓度差)等. 事实上,一种力可以引起多种流,一种流也可以是多种力作用的结果. 因此流与力的线性关系应写成

$$J_i = \sum_j L_{ij} X_j,$$

称 L_{ij} 为昂萨格系数.

美国物理化学家昂萨格(Onsager,L.)在 1931 年利用系统的细致平衡条件证明:存在关系 $L_{ij} = L_{ji}$,即第 i 个力对第 j 个流的影响系数 L_{ij} 等于第 j 个力对第 i 个流的影响系数 L_{ji}. 此关系称为昂萨格倒易关系,它是系统在非平衡线性区普遍存在的关系,与系统的具体性质无关.

最小熵产生原理(principle of minimum entropy production) 线性热力学的基本原理之一. 指近平衡态时系统演化过程中熵变化所遵从的规律,由比利时学者、著名物理化学家、诺贝尔化学奖获得者普利高津(Prigogine,I.)提出并证明. 其内容可概括为:处在非平衡情况下的系统,当离平衡态不太远,其内部存在的各种流与力之间仍能用线性关系表示时,人们称系统处在非平衡线性区;此时系统向着熵增加速度不断减小的方向演化,直至最终达到一个熵增加最小的状态;这样的演化过程及最终达到的状态,不因系统而改变.

对应于最小熵产生的系统定态与平衡态没有本质区别,此时不会形成任何有序结构. 而且若控制条件使系统逐渐接近并达到孤立系统,则此最小熵产生定态连续变为平衡态. 最小熵产生原理与昂萨格倒易关系(参见"昂萨格倒易关系")都与具体系统性质无关,它们构成了非平衡线性区统计物理的两大基石,由此建立起这一领域内的所有理论.

等概率假设(equal-probability hypothesis) 统计物理学的基本假设之一. 对于由大量粒子组成的系统,粒子可能有多种分布,系统也就可以存在多种可能状态. 等概率假设是指对于系统的各种可能状态,若没有其他条件限定,就假定一切可能出现的状态都是等概率的.

孤立系统与外界无任何联系,其能量保持不变. 若用相空间点表示系统的状态,则自然认为在等能面 E 与等能面 $E+\Delta E$ 间各点出现的概率应是一样的,即分布密度函数 ρ 在此范围内为常数,在此范围外为零. 由此推出微正则系综. 进一步分析,得到与外界仅有能量交换和与外界有能量与物质交换的其他两种平衡情况下,系统的分布函数和各种性质,进而建立起了统计物理学.

弛豫(relaxation) 系统从一个定态向另一定态演化的过程. 其物理量随时间变化的规律服从弛豫方程

$$\frac{d\mu(t)}{dt} = \frac{1}{\tau}[\mu(\infty) - \mu(t)],$$

其中,τ 称为驰豫时间,$\mu(\infty)$ 为新稳态值. 弛豫过程中系统状态称为弛豫态.

弛豫时间(relaxation time) 见"弛豫".

涨落(fluctuation) 系统状态变量随机地偏离其平均值的现象. 当巨大系统的宏观控制条件确定

之后，系统的宏观状态平均值不再变化，但由于各个子系统状态仍受到随机扰动，其整体状态变量并非完全不变，而在其平均值附近随机地变动，这种现象称为涨落.

人们将这种物理现象推广到其他统计计算工作. 对于进行多次测量以取得平均数据的事物，其每一次测量值与平均值的差别，也称为涨落. 即使平均值相同，若涨落不同，则反映它们各自随时间变化的偏高程度不同. 涨落小的，其平均值更反映系统的各个时刻的真实状态.

在统计物理学中，系统平衡态涨落的相对值与其所包含子系统个数的平方根成反比. 系统发生相变时，涨落将会被放大，成为巨涨落，而使系统从一个宏观状态变成另一宏观状态.

涨落耗散定理(fluctuation dissipation theorem) 线性热力学的基本定理之一. 描述处在线性近平衡区的系统，其输运系数 L_{ij} 与相应流 J_i 和力 X_j 的热涨落之间关系. 对不同的输运过程其形式可不同. 如线性电路中，电阻与热噪声电动势的均方值成正比，即

$$\overline{(\Delta\varepsilon)^2} = 4\pi RkT.$$

又如布朗运动粒子在时间 t 内的均方位移与扩散系数成正比，即

$$\overline{(\Delta x)^2} = 2Dt.$$

涨落耗散定理可以把某些系统的非平衡性质(涨落)通过平衡性质表示出来.

空间结构(spatial structure) 一类系统有序结构. 指系统状态变量随空间平移出现周期变化的情况，即满足下式：

$$X(\boldsymbol{r},t) = X(\boldsymbol{r}+\boldsymbol{a},t),$$

其中 $\boldsymbol{a}$ 为空间周期. 在非平衡统计物理中，形成空间结构的原因在于系统与外界进行物质、能量的交换，使其远离平衡态. 常见的空间结构有某方向上的周期状态分布，平面上的三角形分布和正方形分布，立体上的圆桶形分布等.

时间结构(time structure) 一类系统有序结构. 指系统状态变量随时间呈现规则变化的情况，例如满足下列周期规律

$$X(\boldsymbol{r},t) = X(\boldsymbol{r},t+T),$$

其中 $\boldsymbol{r}$ 为空间坐标，T 为时间变化周期. 这是在非平衡条件下，由于系统与外界存在物质、能量的变换，系统表现出类似时钟周期变化的一种有序结构.

时空结构(time-spatial structure) 一类系统有序结构. 系统状态变量同时随时间和空间有规律地变化，如呈现类似波动的周期变化的情况，其状态变量满足下式：

$$X(\boldsymbol{r},t) = X(\boldsymbol{r}+\nu\Delta t, t+\Delta t),$$

其中 ν 为波的传播速度. 它也是在非平衡条件下，系统形成的一种有序结构.

序参量(order parameter) 相变理论和协同学的基本概念. 影响系统(特别是复杂系统)演化的因素往往有许多，但起支配作用的可能只有少数几个变量，协同学中把这种对形成系统自组织结构起主导作用的变量称为序参量. 序参量由慢变量、不稳定模担任. 在系统相变过程中，不稳定模支配稳定模，慢变量支配快变量(参见“役使原理”).

役使原理(slaving principle) 亦称支配原理. 协同学的基本原理之一. 协同学创始人哈肯(Haken, H.)推广绝热消去法和中心流形定理而提出的一个系统演化原理. 其内容为：对于某些系统，在一定条件下慢变量成为序参量，支配快弛豫变量，不稳定模支配稳定模，从而主导系统的演化趋势. 役使原理表述了协同学的核心思想，在复杂系统研究中有广泛应用.

快弛豫变量(fast relaxing variable) 描述系统临界行为的一种变量. 某些系统在临界点处的参变量可以分为两种类别：一类数目较多，变化较快，但对系统演化及新结构的形成不起决定性作用，称为快弛豫变量，简称快变量；另一类变量则数目较少，变化较慢，但对系统演化和宏观结构的形成起支配作用，称为慢弛豫变量，简称慢变量. 慢弛豫变量亦称为系统的序参量. 例如对于方程

$$\frac{\mathrm{d}q}{\mathrm{d}t} = -\gamma q + a\exp(-\delta t)$$

(其中 γ 是阻尼系数，$\gamma \gg \delta$)，可通过积分求出其解为

$$q(t) = \frac{a}{\gamma-\delta}[\exp(-\delta t) - \exp(-\gamma t)]$$

$$\approx \frac{a}{\gamma}\exp(-\delta t) \equiv \frac{F(t)}{\gamma}.$$

另外，通过直接令 $\frac{\mathrm{d}q}{\mathrm{d}t}=0$，亦可简捷地求得同样正确的结果. 这里已见到了绝热消去法的基本思路. 再看下面的系统：

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\alpha x - axy, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = -\beta y + bx^2,$$

其中 $\beta>0$，$\beta\gg\alpha$，此式中 x 相当于原来的 F，y 相当于原来的 q. 可以直接令 $\frac{\mathrm{d}y}{\mathrm{d}t}=0$ 而求得

$$y(t) \approx \frac{bx^2(t)}{\beta}.$$

用慢变量 x 表示快变量 y，则有

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\alpha x - \frac{ab}{\beta}x^3,$$

进而可求出稳定定态解

$$x_0 = \begin{cases} 0 & (\alpha > 0), \\ \pm\sqrt{\dfrac{|\alpha|\beta}{ab}} & (\alpha < 0). \end{cases}$$

在上面的处理中，x 起支配作用，描述系统秩序的形成，是慢变量，也称序参量. y 起次要作用，是快变量. 在物理学中，快变量也称稳定模，慢变量也称不稳定模. 于是，慢变量支配快变量也可说成是不稳定模支配稳定模. 在上例中，假设 $\gamma \gg \delta$ 和假设 $\beta \gg \alpha$ 都称为“绝热近似”假设，由这种假设而进行的近似运算称“绝热消去”技术.

慢弛豫变量(slow relaxing variable)　见“快弛豫变量”.

稳定模(stable mode)　见“快弛豫变量”.

不稳定模(unstable mode)　见“快弛豫变量”.

弛豫态(relaxing state)　系统状态变量随时间而变的状态. 与定态相反，弛豫态按系统演化规律形成一条演化轨道，对不同初始条件，系统有不同弛豫态演化轨道.

弛豫时间(relaxing time)　动力学系统的一种特征时间. 系统的某种变量由暂态趋于某种定态所需要的时间. 在统计力学和热力学中，弛豫时间表示系统由不稳定定态趋于某稳定定态所需要的时间. 在协同学中，弛豫时间可以表征快变量的影响程度，弛豫时间短表明快变量容易消去.

熵(entropy)　热力学系统的一种状态函数. 在平衡态热力学中，其改变量等于可逆过程的热温商(热量除以温度之商)的总和. 一般地，熵可表征系统所处状态的无序、混乱或无组织程度. 熵概念最初来自热力学中的可逆过程. 热力学体系经过程 P 由状态 1 变化到状态 2 之后，如果用某种方式能使体系和环境都完全复原而不留下任何变化痕迹，则过程 P 称为可逆的. 反之，该过程为不可逆的. 自然界所发生的大量自发过程都是不可逆的. 热力学中的卡诺循环是可逆过程的例子，它通过等温可逆膨胀、绝热可逆膨胀、等温可逆压缩和绝热可逆压缩四个过程而回到原来状态. 整个过程的热温商之和等于 0，即

$$\frac{Q_{吸}}{T_{高}} + \frac{Q_{放}}{T_{低}} = 0,$$

其中对体系而言 $Q_{吸}$ 为正值，$Q_{放}$ 为负值；$T_{高}$ 和 $T_{低}$ 分别是高温和低温热源的绝对温度.

由卡诺定理可知，任何一个可逆过程都可以看做是由无数个小卡诺循环组成，它们的热温商之和亦为 0. 设体系由平衡态 A 变到平衡态 B，在 A 和 B 之间有各种不同的可逆路径. 可以证明，由不同可逆路径算得的热温商总和都相等，它仅由状态 A 和 B 决定而不依赖于路径. 体系经路径 1 由 A 到 B，再经路径 2 返回 A 是一个可逆循环，故

$$\oint \frac{\delta Q_{R}}{T} = 0,$$

由此式即可得出一个重要结论：体系必然存在一个状态函数，它的改变量等于可逆过程中热温商的总和. 这个状态函数即称为熵，用 S 表示. 注意，熵定义公式的正确写法为

$$\mathrm{d}S \equiv \frac{\delta Q_{R}}{T},$$

这里 $\mathrm{d}S$ 是微小熵变，δQ_{R} 为可逆过程中微小的热量改变(δ 不能写成 d)，T 为体系的绝对温度. 特别指出定义中必须标出 R 以表示可逆过程，这表明无论可逆或不可逆过程的熵变都是通过可逆过程来定义的. 只有通过设计某种可逆过程才能求出过程的熵变量值. 熵定义有时也可写作

$$\Delta S \equiv S_{A} - S_{B} \equiv \int_{A}^{B} \frac{\delta Q_{R}}{T}.$$

热力学第二定律说的是不可能把热从低温物体传到高温物体而不引起其他变化；或者说，不可能从单一热源取出热，使之完全转化为功，而不发生其他变化. 用熵函数的形式表示就是

$$\Delta S - \sum_{i} \frac{\delta Q_{i}}{T_{i}} \geqslant 0,$$

这里左边第一项为过程的熵变，第二项为实际过程中热温商总和. 对于绝热过程 $\delta Q_{i}=0$，上式变为 $\Delta S \geqslant 0$；对于孤立系统，也有 $\Delta S \geqslant 0$.

玻耳兹曼(Boltzmann，L. E.)进而又建立了描述熵与体系微观状态的玻耳兹曼关系式. 它指明体系宏观状态确定时，状态函数熵 S 和总微观状态数 P 之间有关系 $S=f(P)$. 根据熵的容量性质，可推出熵公式的惟一表达式为

$$S = k \ln P,$$

式中 k 为玻耳兹曼常数，等于气体常数 R 除以阿佛加德罗常数 N. 因为式中 S 是宏观物理量，P 是微观物理量，此式沟通了宏观与微观之间、熵与无序度的定量关系. 由此，熵的概念也有了明确的微观解释，即高熵对应无序，低熵对应有序. 自发过程熵增大，最大熵对应于平衡态等.

由微观上看，对应于平衡态的体系最混乱，这通常称为熵混沌. 在统计力学中，还可把熵表示为概率密度对态矢量的积分，从而得到吉布斯熵公式

$$S = -k \int \mathrm{d}\boldsymbol{X}^{N} \rho(\boldsymbol{X}^{N}) \lg[C^{N} \rho(\boldsymbol{X}^{N})],$$

其中 $\boldsymbol{X}^{N}$ 是 $3N$ 个自由度闭合经典系统的态矢量，ρ 为$6N$ 维相空间的概率密度，而 C^{N} 只是归一因子.

对于量子系统，吉布斯熵取更普遍的形式：

$$S = -k \,\mathrm{tr} \hat{\rho} \ln \hat{\rho},$$

其中 $\hat{\rho}$ 是密度算符，迹是对任意一组正交完备基求出的. 对于不可逆过程，通常的做法是把熵变写成两部分之和：

$$\mathrm{d}S = \mathrm{d}_{r}S + \mathrm{d}_{i}S,$$

其中 $\mathrm{d}S$ 表示体系的总熵变，$\mathrm{d}_{r}S$ 表示体系与环境进

行能量或物质的交换引起的体系的熵交换，d_iS 是系统内部的不可逆过程所引起的熵产生，由热力学第二定律知，$d_iS>0$.

序(order)　表示事物关连或组织性结构的一种数学概念. 这里特指由系统演化、对称破缺而形成的时空或功能结构.

无序(disorder)　序的否定概念. 系统在时空之间或功能方面高度对称、无法区分的形态，称为无序. 无序的系统通过与环境进行物质和能量的交换可演化出有序结构，即耗散结构理论研究的内容.

时空关联(space-time correlation)　刻画系统有序性的概念. 系统参量 q 的时空分布的关联函数 $\langle q(x,t),q(x',t')\rangle$ 不为零时，称其具有时空关联. 它可反映涨落的强度和衰减速度、相干运动的持续时间或在空间中的相关长度等.

信息熵(information entropy)　描述信息源各可能事件发生的不确定性的量度. 其表达式为

$$I=-\sum_i P_i\lg P_i,$$

其中 P_i 是该事件诸可能结果的概率.

显然，对各事件发生的先验知识越少，即各 P_i 越近于相同，可以证明其信息熵就越大. 反之，若事件发生是确定性的(即某 $P_i=1$)，则其信息熵为零. 在物理系统中，信息熵也可以看做是对系统状态混乱程度或随机程度的量度. 在热力学中，玻耳兹曼熵定义为

$$S=k\lg N,$$

其中 k 是玻耳兹曼常数，N 是可能的微观状态数目. 若 N 个微观状态的概率相等，均为 $P_i=1/N$，显然玻耳兹曼熵 S 就和信息熵 I 有同样的含义. 对于连续信息源 X，若事件发生于 x 的概率密度为 $p(x)$，则信息熵 $I(x)$ 定义为

$$I(x)=-\int_{-\infty}^{+\infty}p(x)\log_2 p(x)\mathrm{d}x.$$

拓扑熵(topological entropy)　一种广义的熵概念，基于系统拓扑性质对熵概念的推广. 描述系统相空间混合速率及系统复杂度的一种指标. 动力学熵是在信息熵之后由柯尔莫哥洛夫(Колмогоров, A. H.)(1958)和西奈(Sinai, Ya. G.)(1959)提出的，它是一种测度论的熵，称 KS 熵. 实际上，这里的 K 还另指一位俄国数学家克雷洛夫(Крылов, H. C.). 受 KS 熵的影响，阿德勒(Adler, R. L.)、柯恩海姆(Konheim, A. G.)和麦克安德鲁(McAndrew, M. H.)于 1965 年提出拓扑熵概念. KS 熵和拓扑熵都可用来描述确定性系统相空间(包括未定义测度的拓扑空间)的"平均混合速率". 1981 年，罗伯特・肖(Shaw, R.)又对动力学熵赋予信息论的解释，用以说明奇怪吸引子上系统行为的不可预见性. 为了刻画拓扑空间 X 中子空间的混合或搅乱过程，先在 X 上引入一种"划分"或"覆盖"，用 $N(A)$ 记对 X 的 A 划分中子区域的个数. 对于多对一映射 f，会出现几个不同区域映射到同一区域的情况. 这时须考虑 f 的逆映射 f^{-1} 不断作用于划分 A 的情况. 如果 M 是 X 的一个子空间，$f^{-1}(M)$ 有多个区域. 于是，划分 $A\vee f^{-1}(A)$ 的区域的个数 $N(A\vee f^{-1}(A))$ 比 $N(A)$ 大得多. 区域个数的变化，反映了动力系统非线性的程度、复杂性的程度、不可预测性的程度. 这便是拓扑熵的基本含义. 令

$$A^n=A\vee f^{-1}A\vee f^{-2}A\cdots\vee f^{-(n-1)}A=\bigvee_{k=0}^{n-1}f^{-k}A,$$

映射 f 相对于划分 A 的拓扑熵定义为

$$h(f,A)=\lim_{n\to\infty}\frac{1}{n}\ln N(A^n).$$

这样定义的拓扑熵 h 与划分 A 有关，一般取所有划分下的最大 h 值为拓扑熵，即

$$h(f)=\sup_A h(f,A).$$

可以看出，只有划分的区域数 $N(A^n)$ 随 n 指数地增长，才能保证 $h(f)$ 不为 0. 这就使拓扑熵可以用来刻画动力学指数的不稳定性. 拓扑熵的具体计算并不容易，但可以通过划分的加细来逼近 h. 如果划分 $A(n)$ 随 n 增加而加细时，$h(f,A(n))$ 单调不减，则 h 的计算公式可写为

$$h(f)=\lim_{n\to\infty}h(f,A(n)).$$

当拓扑熵 h 大于 0 时，一定存在李-约克意义上的混沌. 对于连续映射 f 来说，若周期点都是 2^k 型的，则拓扑熵 $h(f)=0$. 混沌系统 f 的拓扑熵 $h(f)\geqslant 0$. N 个符号的符号序列空间 $\Sigma(N)$ 上的移位映射 σ 的拓扑熵为 $\ln N$，帐篷映射的拓扑熵为 $\ln 2$.

激光协同学模型(synergetic model for laser)　基于协同学原理建立的激光模型. 即从激光的微观机制出发，根据协同学原理所建立的描述激光现象的宏观序参量方程.

激光系统是光振幅为 b_λ 的光场与偶极矩为 α_μ 的原子之间相互作用而形成的. 它们都满足带有随机力的相互耦合阻尼振动方程，具体形式为

$$\frac{\mathrm{d}b_\lambda}{\mathrm{d}t}+(\tilde{k}+\mathrm{i}\omega)b_\lambda+\mathrm{i}\sum_\mu g_{\mu\lambda}\alpha_\mu=F_\lambda(t),\quad(1)$$

$$\frac{\mathrm{d}\alpha_\mu}{\mathrm{d}t}+(\tilde{\gamma}+\mathrm{i}\nu)\alpha_\mu-\mathrm{i}\sum_\lambda g^*_{\mu\lambda}\sigma_\mu b_\lambda=\Gamma_\mu(t),\quad(2)$$

$$\frac{\mathrm{d}b_\lambda}{\mathrm{d}t}=-\tilde{k}b_\lambda-\mathrm{i}\sum_\mu g_{\mu\lambda}\alpha_\mu+F_\lambda(t),\quad(3)$$

$$\frac{\mathrm{d}\alpha_\mu}{\mathrm{d}t}=-\tilde{\gamma}\alpha_\mu-\mathrm{i}\sum_\lambda g^*_{\mu\lambda}\sigma_\mu b_\lambda+\Gamma_\mu(t),\quad(4)$$

其中，$\tilde{k}$，$\tilde{\gamma}$ 为阻尼系数，ω，ν 为固有频率，$g_{\mu\lambda}$，$g^*_{\mu\lambda}$ 分别为光场与原子磁矩耦合系数，σ_μ 为粒子数反转，

$F(t)$，$\Gamma(t)$为随机力. 粒子数反转受随机力 $\Gamma_{\sigma_\mu}(t)$、与平衡态时粒子数 d_0 的偏离、光场与偶极矩相互作用三者影响，具体为

$$\frac{d\sigma_\mu}{dt}=\gamma_{11}(d_0-\sigma_\mu)+2i\sum_\lambda(g_{\mu\lambda}\alpha_\mu b_\lambda^+-c.c.)+\Gamma_{\sigma_\mu}(t), \quad (5)$$

其中 c.c. 代表复量的共轭. 原子偶极矩 α_μ 弛豫时间远小于光场 b_λ 及反转数 σ_μ 的固有时间. 利用协同学役使原理，在求解 α_μ 时将 b_λ，σ_μ 作为常数，且只考虑单一频率，去掉下角 λ 可得到

$$\alpha_\mu(t)=\frac{i}{\gamma}g_\mu^* b\sigma_\mu+\hat{\Gamma}_\mu(t). \quad (6)$$

同样根据 σ_μ 的弛豫时间小于光场的弛豫时间，可得

$$\sigma_\mu=d_0\left(1-\frac{4g^2bb^+}{\gamma_{11}\gamma}\right). \quad (7)$$

将(6)，(7)式结果代入(2)式，得到光场 b 满足的简化方程

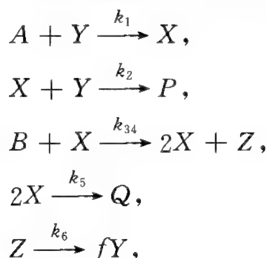
$$\frac{db}{dt}=-\alpha b-\beta b^3+\hat{F}(t), \quad (8)$$

$$\alpha=k-\frac{g_\mu^2}{\gamma}D_0, \quad \beta=\frac{4g_\mu^2k}{\gamma_{11}\gamma},$$

$$\hat{F}(t)=F(t)-\hat{\Gamma}_\mu(t), \quad D_0=\frac{k\gamma}{g^2}.$$

方程(8)为尖点突变方程，它以简捷的方式解决了激光相变的很多问题.

俄勒冈模型(Oregonator model)　化学反应耗散结构的重要模型之一. 对于丙二酸被溴酸氧化的化学震荡反应(参见"BZ 反应")，美国 Oregon 大学的菲尔德(Field，R. J.)、诺伊斯(Noyes，R. M.)等人对其反应机理进行了研究，并提出了不少于 11 步反应组成的详细的动力学机制. 进一步简化为用关键的三种中间产物的物质来表示系统的震荡行为，它们是亚溴酸[$HBRO_2$]、溴离子[Br^-]、铈离子[Ce^{4+}]，分别用 X=[$HBRO_2$]，Y=[Br^-]，Z=[Ce^{4+}]表示. 按照关键的五步反应写出式子：



$$A+Y\xrightarrow{k_1}X,$$
$$X+Y\xrightarrow{k_2}P,$$
$$B+X\xrightarrow{k_{34}}2X+Z,$$
$$2X\xrightarrow{k_5}Q,$$
$$Z\xrightarrow{k_6}fY,$$

其中 k_1，k_2，k_{34}，k_5，k_6 为反应速率常数，f 是一个可调化学计量常数. A，B 为反应物，P，Q 为反应产物. 这样对 X，Y，Z 可列出其反应扩散方程，在空间均匀时，有

$$\begin{cases}\dot{X}=k_1AY-k_2XY+k_{34}BX-k_5X^2,\\ \dot{Y}=-k_1AY-k_2XY+fk_6Z,\\ \dot{Z}=k_{34}BX-k_6Z.\end{cases}$$

经过适当变换，可得到规范的方程

$$\begin{cases}\dot{x}=s(y-xy+x-gx^2),\\ \dot{y}=\frac{1}{s}(-y-xy+fz),\\ \dot{z}=w(x-z).\end{cases}$$

此即称为俄勒冈模型，通过此数学模型，可以解释 BZ 反应出现的各种时空有序结构.

福克尔-普朗克方程(Fokker-Planck equation)　在随机层次上描写粒子概率分布变化的一种方程. 设 f 为概率密度，$K=\gamma q$ 为漂移系数，Q 为扩散系数，则福克尔-普朗克方程为

$$\frac{\partial f}{\partial t}=-\frac{\partial}{\partial q}(\gamma qf)+\frac{1}{2}Q\frac{\partial^2 f}{\partial q^2}.$$

如果假定对应的郎之万方程为

$$\frac{dq_i}{dt}=K_i(q_i)+F_i(t),$$

并且 δ 相关涨落力 F 满足

$$\langle F_i(t)F_j(t')\rangle=Q_{ij}\delta(t-t'),$$

则有下述形式的福克尔-普朗克方程：

$$\frac{\partial f}{\partial t}=-\nabla q\{Kf\}+\frac{1}{2}\sum_{ij}Q_{ij}\frac{\partial^2 f}{\partial q_i\,\partial q_j}.$$

一维福克尔-普朗克方程的定态解为

$$f(q)=N\exp\left(\frac{2}{Q}\int_{q_0}^{q}K(q)dq\right).$$

麦克斯韦妖(Maxwell demon)　一种假想的小精灵. 它处在某容器的分界面上，可将高速粒子驱往一侧而将低速粒子驱往另一侧，从而导致该系统熵自发地减少. 维纳(Wiener，R.)曾形象地解释这种现象是不可能的，因为麦克斯韦妖要判断粒子的速度，需要大量的信息，亦即必须摄入足够的负熵才能完成这一繁重的任务.

李-约克定理(Li-Yorke theorem)　刻画确定性混沌的重要定理之一. 由李天岩和约克(Yorke，J. A.)于 1975 年发表的论文"周期 3 意味着混沌"("Period three implies chaos")提出. 其主要内容为：一维闭区间[0，1]到其自身的连续映射 $x_{n+1}=f(x_n)$若出现周期 3 的循环，亦即对不同点 $a\to b\to c$，第四点 d 必然回到 a 的外侧，包括其特殊情况 $d=a$，亦即确实出现一个周期 3 的点，则这时可出现任何正整数的周期 n.

标度(scaling)　亦称尺度. 用以刻画自然界尺度量级大小关系的概念. 复杂系统中含有大小不同的尺度，其差别可高达好几个量级. 如大气中大的涡漩尺度可达数千公里，而小涡漩的尺度仅有几毫米；又如海岸线，大的弯曲达上千公里，而小的弯曲仅几厘米. 相应的物理量，如涡漩的速度、海岸线的长度等，也都随尺度变化而具有不同的数量级. 不同尺度 r 或 t 的物理量 $f(r)$或$f(t)$存在如下的自相似关

系，通常在统计意义下形如：

$$f(\lambda r)=\lambda^{\alpha}f(r) \text{ 或 } f(\lambda t)=\lambda^{\alpha}f(t),$$

即尺度由 r(或 t)变成 λr(或 λt)后，物理量 $f(\lambda t)$就是原来 $f(t)$的 λ^{α} 倍．上面两个式子称为标度律，α 称为标度指数．

标度律(scaling law)　见"标度"．

标度不变性(scaling invariance)　一种标度对称性．即对于尺度变换的不变性质．例如，海岸线的长度随尺度变化而变化，但是作尺度变换后，海岸线的分数维是不变的；在二阶相变中，自由能随临界温度和磁场而变化，但是二阶相变的海界指数是不变的．又如纳维-斯托克斯方程

$$\frac{\partial \boldsymbol{v}}{\partial t}+\boldsymbol{v}\cdot\nabla\boldsymbol{v}=-\frac{1}{\rho}\nabla p+\nu\Delta\boldsymbol{v}$$

经如下标度变换后，其纳维-斯托克斯方程形式不变：

$$x'=\lambda x,\ t'=\lambda^{1-\alpha}t,\ \boldsymbol{v}'=\lambda^{\alpha}\boldsymbol{v},$$

$$\left(\frac{p}{\rho}\right)'=\lambda^{2\alpha}\left(\frac{p}{\rho}\right),\ \nu'=\lambda^{1+\alpha}\nu.$$

自组织临界性(self-organized criticality)　简称 SOC．动力系统自发地演化到临界态的特性．这个现象由贝克(Bak，P.)、汤超(Tang，C.)和维森弗特(Wiesenfeld，K.)于 1987 年用数值模拟方法首先得到．他们认为，一个具有耗散的动力学系统不需要任何外部参量的控制与调节，系统就能通过自组织行为自发地演化到临界状态．这种状态有别于统计物理中相变时的临界态，是远离平衡的一种吸引子．为清晰地说明 SOC，最典型的实例是沙堆模型．如果在一平面上缓慢播撒沙粒使其聚积成一个沙堆，沙堆的形状虽是粗糙的，但它的特征是由其最大倾斜角来确定的．当系统达到最大倾斜角后，继续播撒沙粒就可能使系统发生大面积的崩塌．崩塌前的位形称为最小稳定位形．它对于任何局域的小扰动是不稳定的，扰动类似于多米诺骨牌，可在所有长度尺度上放大，从而扰动可以导致从单个沙粒的移动到整个沙堆的崩塌．这时系统不存在特征长度与特征时间，是一种典型的临界行为．最小稳定位形与系统的初始状态以及播撒沙粒的方式都无关，而是通过自组织行为达到的一种统计定形．在模拟计算中得到系统的功率谱和崩塌前稳定集团的分布函数都满足幂函数规律，表明该系统具有非线性耗散动力系统的普适行为．目前，科学家们已将 SOC 理论应用到生命科学、地球科学和凝聚态物理中，其中最突出的是用 SOC 理论来解释地震发生的机理．

测度熵(measure entropy)　刻画系统不确定性和复杂性的一种概念．设(Ω,F,P)为给定的概率空间，其中 Ω 为样本空间，F 为事件域，P 为概率测度．(Ω,F,P)上的一个可列划分 φ 是指满足下列条件的一列基本事件 A_i：

1. A_i 属于 $F(i=1,2,3,\cdots)$.
2. $A_i\cap A_j=\varnothing,i\neq j$.　　(1)
3. $\Omega=\bigcup_{i=1}^{\infty}A_i$.

对给定划分可以定义关于 φ 的信息函数

$$I(\varphi,\omega)=-b\sum_{i=1}^{\infty}1_{A_i}(\omega)\lg P(A_i),\qquad(2)$$

其中 $1_{A_i}(\omega)$是 A_i 的示性函数，

$$1_{A_i}(\omega)=\begin{cases}1, & \omega\in A_i,\\ 0, & \omega\notin A_i,\end{cases}\qquad(3)$$

系数 b 是一个可选择的常数．直观上可把 φ 解释为一个物理实验，A_i 为实验的可能结果．$I(\varphi,\omega)$在每一可能的结果上取常值．则该划分 φ 的熵 $H(\varphi)$定义为

$$H(\varphi)=\int_{\Omega}I(\varphi,\omega)P\mathrm{d}\omega$$
$$=-b\sum_{i=1}^{\infty}P(A_i)\lg P(A_i).\qquad(4)$$

信息函数 $I(\varphi,\omega)$在 A_i 上值的大小，反映了事件 A_i 的不确定性，亦即实验中出现 A_i 所包含的信息．熵 $H(\varphi)$则是整体上对实验 φ 的不确定性的整体度量，称为测度熵．

主方程(master equation)　一种动力学方程．协同学、耗散结构等理论中经常采用的一类从随机层次描写系统演化的方程．设随机变量的概率分布函数为 P，它是随机变量 x 和时间 t 的函数，P 的时间变化率可表示为

$$\frac{\mathrm{d}P(x,t)}{\mathrm{d}t}=\text{增加率}-\text{减少率},$$

其中"增加率"由 x'到 x 的转移组成，"减少率"由 x 到 x'的转移组成．将两者具体化，就得出主方程表达式

$$\frac{\mathrm{d}P(x,t)}{\mathrm{d}t}=\sum_{x'}w(x,x')P(x',t)-P(x,t)\sum_{x'}w(x',x).$$

研究主方程的关键问题在于如何显式地确定转移率 w．如果主方程有惟一定态解，并满足细致平衡条件，它的解就可以通过求和或者积分求得．物理学中对于热平衡系统，细致平衡原理一般可由微观可逆性导出．在远离平衡的物理或非物理系统中，细致平衡原理只在特殊情况下成立．因而，求主方程的严格解通常很困难．

适应(adaptation)　亦称自适应．一种系统演化行为机制．系统通过改变结构参数、控制策略等，力求在环境条件发生变化时保持原有功能、继续正常运行的能力或技术．生物体的适应能力包括：生物个体的生理适应，如在各种不同环境条件下保持体温、

血压等生理参数；心理的适应，如对生存环境中群体行为的适应；种族群体的适应，即靠遗传、变异来适应环境变化，属于物种进化范畴；人类对社会变革的适应；以及包括生命系统和技术系统在内的一般系统功能的适应，例如，工作于复杂多变环境下的太空探测器或工业控制系统等.

适应控制的形式包括简单的参数调整、算法调整、结构调整、各种搜索和学习技术等.

定态(steady state)　不随时间改变的系统状态集. 对于由微分方程表示演化规律的系统，当方程的解不随时间变化时，称为定态解. 定态解受扰动时，按其演化规律，若系统运动，并最终恢复到原定态解，则称该定态解是稳定的；反之，则称该定态解是不稳定的.

协同学(synergetics)　一种系统自组织理论. 研究复杂系统中由子系统间相互协同作用而自发地形成宏观尺度上的时间、空间或功能的有序结构，即系统的总体宏观效应. 它由德国学者哈肯(Haken, H.)于 1977 年提出. 协同学的思想最初来自物理学、化学中的一些自组织现象，如贝纳德流和泰勒流中宏观花纹的形成、激光形成的相干振荡机制、贝洛索夫-扎波金斯基反应出现的时-空振荡结构等. 哈肯认为，形成这些自组织现象最基本的原理是役使原理(参见“役使原理”)，它指出维数极大的巨系统中大量快变的运动模式是由少量慢变的模式来支配的，而后者决定了系统的较为稳定的宏观行为. 协同学问题的数学模型是一种时间演化方程

$$\dot{q}=N(q,\alpha)+F(t),$$

其中 q 为时空状态，一般地是空间坐标 x 和时间 t 的函数，N 为非线性函数，α 为控制参量，即外界施加的调控参数，$F(t)$表示随机涨落的影响. 需要研究的是这种系统的旧运动形态如何由于参数 α 的改变和随机涨落的影响而失去稳定，并由新的稳态来代替. 具体步骤包括：在原参数 α_0 下求出系统的某稳态(包括周期态等)解 $q_0(t)$；在 $q_0(t)$邻近进行线性化，令 $w=q-q_0$，并得到线性近似方程

$$\dot{w}=Lw,$$

从而得到稳定模 $v_s(t)$和不稳定模 $v_u(t)$；运用役使原理将被支配的稳定模用不稳定模的序参量 ξ_u 来表示，并最后得到比较简单的序参量方程

$$\dot{\xi}_u=M(\xi_u,F),$$

由此即可得到维数大为降低的宏观运动模态.

与通常控制系统中按既定目标设计的控制作用不同，哈肯强调这里的控制参数 α 只代表一些条件(如外加能量)而并未提供多少信息，新结构的形成是由系统内部的相互作用以及随机起伏的选择而进行的，因此称为自组织过程. 协同学原理和方法进而被用于解释、描述许多自然和社会现象，乃至人类认知过程，从而为人们对各种复杂系统的理解、设计、管理和调控提供了有力的指导.

洛特卡-沃尔泰拉方程(Lotka-Volterra equation)　亦称捕食者-被捕食者模型. 反映捕食者-被捕者系统演化特点的一组方程. 对于一个由捕食者(Y)与被捕食者(X)组成的系统，它们的演化行为具有下列特点：

1. 被捕食者获取食物自我发展：$A+X\rightarrow 2X$.

2. 捕食者消灭被捕食者，促进自身发展：

$$X+Y\rightarrow 2Y.$$

3. 捕食者自身消亡：$Y+D\rightarrow D$.

进一步可建立它们演化的数学模型：

$$\begin{cases}\dot{X}=AX-XY,\\ \dot{Y}=XY-BY.\end{cases}$$

此模型存在一组定态解$\begin{pmatrix}B\\A\end{pmatrix}$. 在微分方程定性理论中，此定态解属中心型，当给定条件使系统偏离此定态解时，系统以偏离值为振幅进行周期振荡. 此模型是由意大利的洛特卡(Lotka, A. J.)在研究地中海鱼类资源发现其周期变化而总结建立的，后由美国的沃尔泰拉(Volterra, V.)进一步整理成上述规范形式. 它不仅能很好反映多数捕食者与被捕食者系统的演化特点，而且方程比较简单，便于研究.

自组织系统(self-organizing system)　能自行演化或改进其组织行为结构的一类系统. 在一般系统论中，其最广泛的含义是指：该系统能在与环境相互作用条件下，通过自身的演化而形成新的结构和功能的系统. 在自然界中，远离平衡态的开放系统形成序结构的自组织过程已在耗散结构和协同学等理论中得到了充分的研究. 在人造的工程系统中，还可能通过精心设计的结构和法则，使系统按不断变化的环境的要求，不断改进自己的控制功能以保持满意的品质，故亦称为自组织控制系统.

自组织的概念最早由阿什贝(Ashby, W. R.)于 1962 年提出，他同时还指出系统组织有序化的前提是其环境有序，并可提供关于序的信息. 自组织系统的基本原理包括：

1. 系统 A 在环境 E 中是自组织系统，则 A 的不定度 $I(A)$不小于环境的不定度 $I(E)$：$I(A)\geqslant I(E)$，即系统有足够的变化余地来适应环境变化.

2. 随着系统的组织化程度提高，其冗余度 R 必随时间增长：

$$\frac{\mathrm{d}R}{\mathrm{d}t}>0,$$

由于冗余度

$$R=-\frac{H}{H_{max}},$$

这表明当 $H_{max}=$常量时有

$$\frac{dH}{dt} < 0,$$

亦即其熵必然在减少. 对于自然界中极为丰富多彩的自组织现象,目前人们能够模仿、实现的只是其中十分简单的一小部分. 学习系统、自适应控制系统等实际上也是自组织的一些具体形式,在系统发生故障时具有重构能力的容错系统也是典型的例子. 从技术实现的角度又可把自组织控制系统定义为:若系统及其环境的先验不确定性可以在系统运行演化过程中通过实时地积累观测信息和施加控制作用而加以减弱,从而使系统能不断改变自己的品质,则它就是一个自组织控制系统.

郎之万方程(Langevin equation) 一类随机微分方程. 与主方程、福克尔-普朗克方程共同组成随机层次描述系统演化的三类方程. 其基本形式为

$$\frac{db}{dt} = -i w_0 b - \kappa b + F(t),$$

其中 b 为随机变量,κ 为耗散强度,$F(t)$ 为随机力,$i = \sqrt{-1}$. 有时也将下面的方程称为郎之万方程:

$$\frac{dq}{dt} = -\alpha q - \beta q^3 + \gamma \Delta q + F.$$

此方程的特点是含有一个线性项 $-\alpha q$,在一定的"阈值"处,系数 α 改变符号;含有一个非线性项 $-\beta q^3$,它对系统的稳定性起重要作用;含有一个"扩散项" $\gamma\Delta$,Δ 是拉普拉斯算子.

反应扩散方程(response-diffusion equation) 一类特定的抛物型偏微分方程. 其形如

$$\frac{\partial u(x,t)}{\partial t} = f(x,u,\nabla u) + D(x,u)\Delta u, \quad (1)$$

其中 x,t 为通常的空间变量和时间变量,∇ 为梯度算子,Δ 为拉普拉斯算子

$$\left(\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right),$$

$u = (u_1, u_2, \cdots, u_m)$ 为 m 维矢量,D 为 $m \times m$ 维矩阵,称为扩散系数. f 为 m 维函数族. 写成分量形式为

$$\frac{\partial u_i(x,t)}{\partial t} = f_i(x,u,\nabla u) + \sum_{j=1}^{m} D_{ij}(x,u)\Delta u_j, \quad (2)$$

它是描写系统演化的一般方程,在许多自然科学领域得到大量使用. 当变量在空间中均匀时,$u = u(t)$,反应扩散方程简化为常微分方程

$$\frac{\partial u}{\partial t} = f(x,u,\nabla u). \quad (3)$$

当系统演化过程中无反应时,$f = 0$,反应扩散方程简化为简单的扩散方程

$$\frac{\partial u}{\partial t} = D(x,u)\Delta u. \quad (4)$$

上述方程有很强的实际背景,在物理学、化学、生物化学、生态学、神经学、系统科学等领域中,对很多实际问题的研究最终都归结为求解反应扩散方程. 通常研究的基本问题是:

1. 以上方程组行波解存在惟一性及稳定性.

2. 初值、边值问题的特解、整体解形式,存在惟一性、稳定性.

3. 定态解的存在性、表达式、解随参数变化的稳定性变化及解分岔结构问题.

4. 求解问题的新计算方法.

贝洛索夫-扎波金斯基反应(Belousov-Zhabotinsky reaction) 简称 BZ 反应. 一种化学自组织现象. 1958 年由两位苏联生物化学家发现而命名. 在进行铈离子催化下柠檬酸或丙二酸的溴酸氧化反应时,当其中各种物质浓度控制于接近平衡态的比例并在均匀边界条件下,生成物呈现均匀混合分布的无序态. 但若改变其浓度比例使之远离平衡态,可以观察到某些组分的浓度会随时间周期变化,并呈现规则的彩色的空间结构,称之为化学振荡或化学钟. 这表明,原来的对称性发生破缺并形成宏观的、时空的序结构. 这也是非平衡系统自组织或耗散结构的一个著名的例子(参见"耗散结构理论").

BZ 反应(BZ reaction) 贝洛索夫-扎波金斯基反应的简称.

社会舆论形成(formation of social opinion) 分析社会舆论发展形成及其稳定特点的一种数学模型. 它由德国学者魏德利希(Weidlish, W.)和哈格(Hagg, G.)利用协同学原理建立起来的.

假定社会总人数为 $2n$,对某一事物每一个人的态度只有两种:同意、反对,其人数分别为 n_1, n_2. 用单一变量 $n = n_1 - n_2$ 表示确定人群舆论的状态. 用 t 时刻状态取 n 值的概率 $p(n,t)$ 作为模型中变量描述系统. 按协同学理论建立 $p(n,t)$ 遵从的方程

$$\frac{\partial p(n,t)}{\partial t} = \sum_{n} \{W(n' \to n)p(n',t) - W(n \to n')p(n,t)\}, \quad (1)$$

其中 $W(n' \to n)$ 表示从 n' 状态向 n 状态的转移概率. 进一步假定:

1. 每一次只有一个人转变观点.

2. 转移概率等于一个人态度转变的可能性与可以进行态度转变的人数的乘积,方程可简化为

$$\begin{aligned}\frac{\partial p(n,t)}{\partial t} = &(n_1 + 1)p\downarrow(n+1,\delta,k)p(n+1,t) \\ &+ (n_2+1)p\uparrow(n-1,\delta,k)p(n-1,t) \\ &- n_1 p\downarrow(n,\delta,k)p(n,t) \\ &- n_2 p\uparrow(n,\delta,k)p(n,t),\end{aligned} \quad (2)$$

其中,$p\downarrow(n,\delta,k)$ 表示处在 n 状态、个人偏好参数为 δ、个人顺从因子取 k 时,一个人向 $n-1$ 状态转变的可能性,$p\uparrow(n,\delta,k)$ 表示同样情况下向 $n+1$ 状态转变的可能性. $\delta > 0$ 表示个人偏好同意观点;$\delta < 0$ 表示偏好反对观点;$k > 0$ 表示个人偏向顺从其

他人意见，$k<0$ 表示个人偏向反对多数人意见.

对方程式(2)取和，得到平均值$\langle n\rangle$满足的方程，同时简化表达式，将对函数取平均值简化为对变量 n 取平均值，再采用新变量 $x=\langle n\rangle/N$，即可得到个人转变态度可能性的表达式：

$$\begin{cases} p\uparrow(n,\delta,k)=\nu\exp[kx+\delta], \\ p\downarrow(n,\delta,k)=\nu\exp[-(kx+\delta)]. \end{cases} \quad (3)$$

总之可得到

$$\frac{\mathrm{d}x}{\mathrm{d}t}=2\nu[\mathrm{sh}(kx+\delta)-x\mathrm{ch}(kx+\delta)], \quad (4)$$

即社会舆论形成模型. 给定参数数值 k,δ,ν，及初条件，求解可得到与实际相符的舆论形成过程的结果. 分析参数变化时对方程定态解的影响，可研究各种情况下舆论形成的不同形式.

超循环理论(hypercycle theory)　一种分子自组织理论. 描述从无生命到生命的进化过程的模型和数学理论. 它是由德国生物物理化学家艾根(Eigen,M.)于 1971 年正式提出的. 按照这种理论，在生命起源和进化过程中，在公认的化学进化和生物学进化阶段之间还存在一个分子自组织阶段，即从生物大分子到原生细胞的进化阶段. 生化过程中存在不同等级的循环，即反应循环、催化循环以及通过循环联系把自催化或自复制单元连结起来的超循环. 超循环代表一种更高的组织形式，艾根为生物大分子的进化提出了一种数学模型：

$$\dot{x}_i=(A_iQ_i-D_i)x_i+\sum_{k\neq i}w_{ik}x_k+\Phi_i,$$

其中，下标 $i=1,2,\cdots,n$ 指明不同的分子种类，即各种可辨别的、带有其特殊遗传信息的自复制分子单元，x_i 是其第 i 种的种群量，$A_iQ_ix_i$ 表示种 i 以自己为模板的合成速率，$-Dx_i$ 为分解速率，$Q_i\in[0,1]$ 表示种 i 正确复制的比例，$w_{ik}x_k$ 表示由种 k 错拷贝而产生种 i 的速率. 种 i 的流之比率为

$$\Phi_i=\frac{\Phi_t x_i}{\sum_k x_k},$$

并可经调节总流量 Φ_t 而使总量

$$\sum_k x_k=\text{常数}.$$

对二元超循环情形，可以具体求出该方程的三个不动点及其本征值，从而分别得到两个种群相互竞争，其中一个被选择，或者产生稳定的超循环组织的结果. 这一结论可以推广到任意 n 元超循环的情形，并成为超循环理论的主要数学结果. 超循环理论成功地把数学方法用于生命起源过程研究，也为复杂系统的自组织理论提供了又一支柱. 当然这里的数学模型还过于简化，还有待实验证实和更深入的研究.

混沌与分形

混沌(chaos)　描述紊乱现象的一种数学概念. 指确定性非线性系统可能产生的一类貌似随机的非周期定态. 混沌和噪声不同，后者是由外部随机驱动力所引起的非周期形态，前者则是由系统内部的非线性机制所决定的，所以混沌也可以说成是内在随机性. 许多科学家相信，自然界的许多复杂现象，如气候、湍流、电子噪声、DNA 序列等可用确定性系统描述的对象，均具有貌似混沌的外在特征. 混沌有敏感初条件的特性，即初始条件的微小差别，将使相空间的轨道迅速分离，所以在混沌系统中短期的时间演化是可以预测的，但长期轨道上的确切位置原则上是不能预测的. 应该指出，混沌的发现在确定论和随机论之间架起了桥梁，是人们对客观世界复杂性更深入的认识.

分形(fractal)　描述复杂的几何形体结构的一种数学概念. 指维数一般取非整数的几何形体. 该术语由曼德勃罗特(Mandelbrot,B.B.)于 1975 年创立，是拉丁文 fractus(破坏)、英语 fracture(断裂)和 fraction(分数)三词综合而成的一个新词. 分形表示一种具有缩放对称性，即自相似性的几何对象，它们在标度变换下具有不变性，从而表示该几何体没有

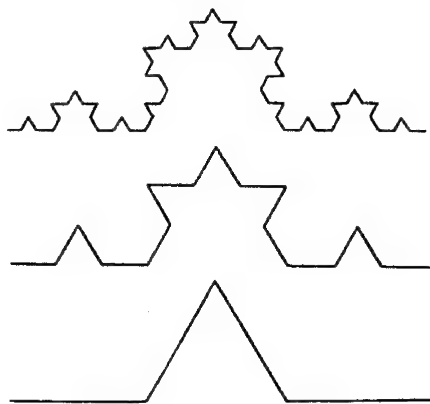


图 1　寇赫曲线

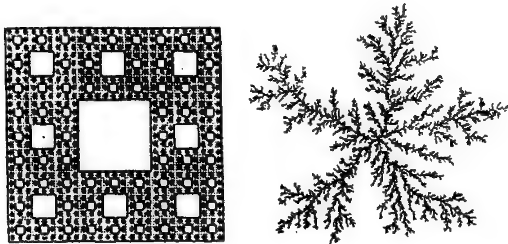


图 2　谢尔品斯基地毯　图 3　二维扩散置限凝聚

特征尺度. 从传统几何学来看，这类几何体是病态

的、不规则的和复杂的. 根据其缺损或稀疏程度,可以导出分形的另一特点,即它的空间维数可以不必是整数而是分数,即分维数 D,它是对直观的欧几里得维数 d 的一种推广.

分形的自然界原型包括雪花、海岸线、海绵、叶脉和毛细血管等. 图 1、图 2、图 3 分别给出了几种典型的分形,图 1、图 2 是规则分形,分别是寇赫曲线与谢尔品斯基地毯,图 3 为二维扩散置限凝聚. 可求得它们的分维数 D 分别为:

寇赫曲线: $D=\frac{\lg 4}{\lg 3}\approx 1.2618$;

谢尔品斯基地毯: $D=\frac{\lg 3}{\lg 2}\approx 1.5849$;

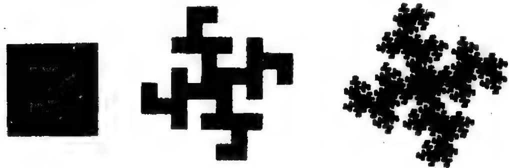
二维扩散置限凝聚: $D\approx 1.66\sim 1.71$.

有多种定义和计算分维数的方法,如盒计数分维数、豪斯多夫分维数、相关函数分维数等. 分形作为一种新的几何概念,现已在许多领域中得到了广泛应用.

分维(fractal dimension) 亦称分形维. 描述几何形体复杂度的一种数学概念. 指数值可取为分数的维数概念. 它是反映该对象所包含点集数量或疏密程度的一种定量度量,从而为某些复杂几何对象(粗糙、破碎和不规则的几何对象,如分形)提供了其他方法不能明确定义的一些性质的测度.

曼德勃罗特(Mandelbrot,B. B.)在其奠基性著作中考察了英国海岸线有多长的问题. 这个问题看似简单,实际上并不容易回答,因为其答案依赖于测量时所使用的尺度单位. 当人们用公里作测量单位时,千米以内的曲折将被忽略;同样,用米或更小的度量单位时,更小的曲折亦不能被反映出来. 在实际测量中,海岸线长度测量单位不能太大(如光年),亦不能太小(如分子级尺度),在这两个自然限度之间存在着若干个数量级的区域,在该区域内,实际测量结果将随测量单位的缩小而变大. 显然,长度不是海岸线等分形曲线长短或曲折程度恰当的定量特征,而这种用普通概念和方法无法进行度量的区域称为"无标度区".

为了更好地理解上面的说明,下面考察数学家



寇赫(Koch, H. Von)构造的寇赫岛中的"海岸线"问题. 将边长为 1 的正方形按上图的形式进行变换并保持面积不变,则一个正方形变成 36 个边长为 1/6 的正方块. 若将谢尔品斯基地毯中每个正方形再按图中的方式变换下去(参见"分形"),则所得到的图形中的正方块的数目将更多,其"海岸线"(总边长)亦越长. 不断变换下去,则得到了寇赫岛,它的海岸线将变得无限曲折,而其边长为无穷. 对于寇赫岛这样的几何对象,长度不能说明其任何性质,而分数维数则是它的更确切的特征量. 按照德国数学家豪斯多夫(Hausdorff,F.)的维数定义,由于测量尺度每缩小 6 倍(边长每经一次变换缩小 6 倍),寇赫岛的海岸线长度增加了 18 倍,因此其豪斯多夫维数为

$$D_0=\lim_{n\to\infty}\frac{\ln N(\varepsilon)}{\ln 1/\varepsilon}=\frac{\ln 18}{\ln 6}\approx 1.613147.$$

显然 D_0 要大于普通的曲线维数 1.

早在 1919 年,数学家豪斯多夫曾提出了分形维数的概念,但这个概念直到 20 世纪 70 年代才开始受到物理学家的重视,并被用来处理无标度性问题. 近 20 年来,随着复杂现象研究的不断深入,各种形式的分数维数得以定义并被深入探讨. 目前已有严格定义的分数维数,包括豪斯多夫维 D_0、信息维 D_1、关联维 D_2、一般分维 D_q(q 为整数)、李亚普诺夫维等. 它们可统一定义如下:用 N 个直径为 ε 的小球覆盖点集 X,令 P_i 记一个点落在第 i 个 ε 球内的概率,若存在极限

$$D_q=\lim_{\varepsilon\to 0}\frac{1}{q-1}\cdot\frac{\ln\sum_{i=1}^{N}P_i^q}{\ln\varepsilon},$$

则 D_q 称为 X 的一般分数维数.

可以证明,当 $q=0,1,2$ 时,D_q 分别为 D_0, D_1, D_2. 进一步地,D_q 系列的分数维数满足不等式

$$d\leqslant D_2\leqslant D_1\leqslant D_0, \tag{1}$$

$$D_q\leqslant D_{q'},\text{若 } q>q', \tag{2}$$

式中 d 为相应的拓扑维数. 根据式(1),分维可定性地定义为一几何对象的维数 D 是分维,若它满足条件 $d\leqslant D$,式中 d 为相应的拓扑维.

从实验数据中提取和计算分维是分形几何理论的近期发展. 通过对分维的计算,人们有可能从少数数据中获得许多复杂现象中的某些规律性结果,并开展对其演化运动的建模与分析.

容量维(capacity dimension) 亦称容积维或柯尔莫哥洛夫维. 一种分数维数. 它与豪斯多夫维数相近,有不同定义方法. 给定一个点集 x,$N(\varepsilon)$是能够覆盖 x 的直径为 ε 的小球的最小数目,如果存在有限极限

$$D_c=\lim_{\varepsilon\to 0}\inf\frac{\ln N(\varepsilon)}{\ln(1/\varepsilon)}, \tag{1}$$

则 D_c 称为 x 的容量维. 与豪斯多夫维数定义相比较,容量维定义的式子中仅多了一项 inf,而在大多数实际计算中,两者是相同的. 基于此项事实,有时它们甚至被认为是一个维数,即豪斯多夫维数. 与容量维类似的维数定义,还有豪克斯(Hawks,J.)于

1974 年引入的上熵维数，其定义是将式(1)中的 inf 改为 sup，即

$$D_H = \lim_{\varepsilon\to 0} \sup \frac{\ln N(\varepsilon)}{\ln(1/\varepsilon)}. \quad (2)$$

豪克斯证明了对于大多数情形，$D_H = D_c = D_0$. 柯尔莫哥洛夫(Колмогоров, А. Н.)于 1959 年曾提出过另一形式相近的分维，其定义是：给定一个点集 x，记 $M(P)$ 为满足下述条件的点集 S 的最大(点的)数目：对于任意 $x, y\in S$，有 $d(x,y)>2P$，若存在有限极限

$$D_K = \lim_{P\to 0} \inf \frac{M(P)}{\ln(1/P)},$$

则 D_K 为集 x 的另一形式的分维. D_K 与 D_0(豪斯多夫维数)相差不大. 实际上，对于线段上的点集而言，$M(P)=N(P)$，因此 $D_K=D_0$.

由于上面定义的三种分数维数均与豪斯多夫维数相差不大，因此，一般把它们视为豪斯多夫维数的同类，有些文献甚至把它们视为相同. 一般地，它们之间存在下述关系：

$D_c \leqslant D_0 \leqslant D_H$,

$D_0 = D_K$，对于线段上的点集 x.

容积维(capacity dimension)　即“容量维”.

柯尔莫哥洛夫维(Kolmogorov dimension)　见“容量维”.

盒计数维(box counting dimension)　见“容量维”.

信息维(information dimension)　分维的一种. 引入信息维数是为了更好地反映分形点集内部的不均匀性，其定义如下：用 $N(\varepsilon)$ 个直径为 ε 的小球覆盖点集 x，令 P_i 为一个点落在第 i 个 ε 球内的概率，如果有限极限

$$D_1 = \lim_{\varepsilon\to 0} \frac{\ln\left[-\sum_{i=1}^{N(\varepsilon)} P_i \ln P_i\right]}{\ln(1/\varepsilon)} \quad (1)$$

存在，则称 D_1 为 x 的信息维. 式(1)等式右边分子括号中的表达式

$$-\sum_{i=1}^{N(\varepsilon)} P_i \ln P_i$$

是著名的信息量公式，这就是信息维名称的来源. 假设每个小球包含点的概率相同，即 $P_i = 1/N$，则信息维还原为豪斯多夫维数. 由于信息维能区分不同小球覆盖点的多少，反映点集内部的不均匀性，因此它是豪斯多夫维数的一种改进和补充. 可以证明，信息维数与豪斯多夫维数的关系是 $D_0 \geqslant D_1$.

关联维(correlation dimension)　分维的一种. 其定义如下：设 $x_1, x_2, \cdots, x_n$ 是系统的一个解序列，令

$$\theta(\varepsilon - |x_i - x_j|) = \begin{cases} 1 & (\varepsilon - |x_i - x_j| \geqslant 0), \\ 0 & (\varepsilon - |x_i - x_j| < 0), \end{cases}$$

$$C(\varepsilon) = \frac{1}{N^2} \sum_{i,j} \theta(\varepsilon - |x_i - x_j|),$$

如果有限极限

$$D_2 = \lim_{\varepsilon\to 0} \frac{\ln C(\varepsilon)}{\ln(1/\varepsilon)}$$

存在，则 D_2 称为系统的关联维. 它极便于进行数值计算.

目前，实践中便于测量的分维只有 D_2. 从上面的定义可以看出，测量 D_2 不需要事先知道系统的背景相空间的维数，也不需要数据序列 $x_1, x_2, \cdots, x_n$ 完全跟踪系统运动轨道. 上述特点在实际系统分析时极其重要. 为说明简单起见，现考虑已测得的简单的一维时间序列 $x_1, x_2, \cdots, x_n$. 把这些数据重新处理，形成一条 m 维相空间的轨道. 用数据支起一高维空间的方法很多，例如取 $m=4$，将 x_1, x_2, x_3, x_4 作为四维空间中的一个矢量 y_1，然后右移一位，把 x_2, x_3, x_4, x_5 作为四维空间的第二个矢量 y_2，以此类推，可构造出四维空间上的运动轨道 $y_1, y_2, \cdots, y_k$. 根据数据 $\{y_i\}$ 及定义，可以对选定的 ε 计算出 D_2. 对于实际分析情形，D_2 的计算结果与 ε 的取值有关(这与理论上 $\varepsilon\to 0$ 时确定 D_2 的情形不同). 一般地，对于足够大的嵌入空间维数 m，当 ε 较大时，$D_2=0$；当 ε 较小时，由于噪声效应，$D_2=m$；当 ε 合适时，D_2 较好地反映了数据的关联维数. D_2 的测量方法较为简单，因此，越来越多的人开始测量自己所研究的对象的维数. 例如，有人用脑电波数据测量相应的分维数，结果发现，当被测者闭目静坐时，$D_2=2.4$；而被测者从数字 300 向前默念时，D_2 上升到 3.0. 虽然分数维对系统描述的意义远未阐明，但众多实验测量出的分维不高的事实可能暗示，完全描述这些复杂现象的模型可能并不需要很多变量. 已证明分数维数 D_0, D_1, D_2 满足下述关系：

$$d \leqslant D_2 \leqslant D_1 \leqslant D_0,$$

式中 d 为拓扑维，D_0 为豪斯多夫维，D_1 为信息维.

马蹄映射(horse shoe map)　亦称斯梅尔马蹄. 一种具马蹄形变换曲线的二维映射. 由美国数学家斯梅尔(Smale, S.)于 1960 年提出，是微分动力系统早期研究中一个里程碑，深化了人们对微分动力系统的通有性、结构稳定性、Ω 集的构造等这些基本概念的认识. 20 世纪 70 年代末以来，马蹄映射更以其作为混沌现象的为数不多的典型数学模型而著称. 以下简述在二维球面 S^2 上构造马蹄映射的一种方法. 设 $f: S^2\to S^2$ 将极点作为不动源点，并以上半开球面为其非稳定流形. 下半闭球面视为一平面上的矩形，作如下变换：记 $f=g\circ l$，其中 l 是线性压缩与拉伸. 于是可以证明，微分同胚 f 的非游荡集 $\Omega=$

$\Omega_1 \cup \Omega_2 \cup \Omega_3$，其中 Ω_1，Ω_3 分别是北极点（不动源点）和另一个不动汇点，Ω_2 则是具有康托尔集结构的 f 的双曲不变集. f 在 Ω_2 上共轭于符号动力系统，于是周期点稠密并存在稠密的周期轨道. 还可证明，以上定义的马蹄映射 f 在 S^2 上是结构稳定的. 在高维流形上亦可定义马蹄映射并证明相应的性质. 马蹄映射的动力学性质具有通常所说的混沌现象的各种典型特征，诸如同时具有整体的结构稳定性和轨道对初条件的敏感依赖等，因而受到广泛的重视.

斯梅尔马蹄（Smale horse shoe） 即"马蹄映射".

厄农映射（Henon map） 在二维空间产生混沌的一种迭代映射. 由法国天文学家厄农（Henon，M.）于 1976 年提出. 其具体形式为

$$x_{n+1} = y_n + 1 - a x_n^2,\ \ y_{n+1} = b x_n.$$

1979 年，他又提出另一形式为

$$x_{n+1} = \cos \alpha x_n - \sin \alpha (y_n - x_n^2),$$
$$y_{n+1} = \sin \alpha x_n + \cos \alpha (y_n - x_n^2).$$

这些映射可以显示保守系统中出现混沌的有趣现象.

豪斯多夫测度（Hausdorff measure） 由豪斯多夫（Hausdorff，F.）提出和命名的一种测度. 为了定量地描述非整数维，豪斯多夫于 1919 年从测量的角度引进了豪斯多夫测度. 该测度是对长度、面积和体积等的推广，也是勒贝格测度的推广. 设在 n 维欧氏空间 $\mathbb{R}^n$ 中对子集 U 定义其直径为

$$|U| = \sup\{ \| x - y \| \mid x, y \in U \},$$

如果子集族 $\{U_i\}$ 为可数个直径不超过 δ 的集构成的覆盖 F 的集类，即

$$F \subset \bigcup_{i=1}^{\infty} U_i,$$

且对每一个 i 都有 $0 < |U_i| \leqslant \delta$，则称 $\{U_i\}$ 是 F 的一个 δ 覆盖. 设 F 是 $\mathbb{R}^n$ 中的任一子集，S 为一非负数，对任意 $\delta > 0$，定义

$$H_\delta^S(F) = \inf\left\{ \sum_{i=1}^{\infty} |U_i|^S : \{U_i\}\ 为\ F\ 的\ \delta\ 覆盖 \right\}. \quad (1)$$

考察所有直径不超过 δ 的 F 的覆盖，并让这些直径的 S 次幂的和达到最小. 当 $\delta \to 0$ 时，$H_\delta^S(F)$ 趋于一极限值，可写为

$$H^S(F) = \lim_{\delta \to 0} H_\delta^S(F). \quad (2)$$

$H^S(F)$ 称为 F 的 S 维豪斯多夫测度. 通常测度只是赋予集以数值"大小"的一种方式，如果集是以合理的方式分解为有限或可数个部分，则整体的数值应该是所有部分数值之和. 可以证明，对于空集 $\varnothing$，有 $H^S(\varnothing) = 0$. 如果 E 包含于 F 内，则 $H^S(E) \leqslant H^S(F)$. 豪斯多夫测度具有平移不变性与旋转不变性. 长度、面积和体积具有众所周知的比例性质，即当比例放大 λ 倍时，曲线的长度放大 λ 倍，平面区域的面积则放大 λ^2 倍，而三维物体的体积则放大 λ^3 倍. 由此可预料，S 维豪斯多夫测度的放大倍数为 λ^S. 其数学表达式为：若 $F \subset \mathbb{R}^n$，$\lambda > 0$，则

$$H^S(\lambda F) = \lambda^S \cdot H^S(F). \quad (3)$$

豪斯多夫维（Hausdorff dimension） 对几何形体的容量进行度量的一种数学描述. 它的数学基础是利用豪斯多夫测度. 设对给定的 F 和 $\delta < 1$，若 $T > S$ 时可推得

$$\sum_i |U_i|^T \leqslant \delta^{T-S} \sum_i |U_i|^S, \quad (1)$$

其中 $\{U_i\}$ 是 F 的一个 δ 覆盖（参见"豪斯多夫测度"）. 对所有的 δ 覆盖取下确界得到

$$H_\delta^T(F) \leqslant \delta^{T-S} \cdot H_\delta^S(F). \quad (2)$$

令 $\delta \to 0$，并设 $H^S(F)$ 为有限值，则由式(2)可得

$$H^T(F) = \begin{cases} 0 & (T > S), \\ H^S(F)\ 为有限值 & (T = S), \\ \infty & (T < S). \end{cases}$$

这说明 $H^S(F)$ 关于不同的 S，存在一个使 $H^S(F)$ 从 ∞ 跳跃到 0 的惟一临界值 S_0，该值就称为 F 的豪斯多夫维数，记为 $D_{\mathrm{H}}(F)$. 通常表示为

$$\begin{cases} H^S(F) = \infty & (0 \leqslant S < D_{\mathrm{H}}(F)), \\ 0 < H^S(F) < \infty & (S = D_{\mathrm{H}}(F)), \\ H^S(F) = 0 & (D_{\mathrm{H}}(F) < S < \infty), \end{cases}$$

由此可见，对于任何一个有确定维数的几何体，若用与它相同维数的"尺"去测量，则可得到一个确定的数值；若用低于它维数的"尺"去量它，结果是无穷大；若用高于它维数的"尺"去测量，结果为零. 所以豪斯多夫维数可对各类几何体（规则的、光滑的、不规则的、不平整的）进行定量的描述. 但按定义直接计算豪斯多夫维数是相当困难的，因为要计算集 F 的维数必须考虑 F 的全部 δ 覆盖，然后从其中选出最好的覆盖，即下确界；另外，还要估计出 S 幂时测度 $H^S(F)$ 的正的有限值，即它的上下界. 这些精确值都是极难求出的. 因此，豪斯多夫维数的实际应用仍具有相当的难度.

分形布朗运动（fractial Brownian motion） 简称 FBM. 一种推广的布朗运动. 它将原布朗运动自相似性质推广为自仿射性. 对一个在一维空间内作布朗运动的微粒，记录它在不同时刻的位移分量 $X(t)$，在 $\Delta t = t - t_0$ 时刻内，$X(t)$ 的改变量为

$$\Delta X(\Delta t) = X(t) - X(t_0),$$

其均方值为

$$\langle \Delta X^2(\Delta t) \rangle \propto \Delta t^{2H} = 2D \cdot \Delta t,$$

其中，H 是一个参量，对于布朗运动 $H = 1/2$，而 D 是扩散系数. 如果将观察时间作一标度变换，令 $\tau = b \cdot \Delta t$，则有

$$\langle \Delta X^2(b\Delta t) \rangle = 2Db \cdot \Delta t = b \cdot \langle \Delta X^2(\Delta t) \rangle.$$

从上式可看出，布朗运动是在标度变换下各向同性地保持着不变，所以它是一个自相似分形. 曼德勃罗特(Mandelbrot,B. B.)推广了布朗运动的概念，他把参量 H 的值定义为 $0<H<1$，而 $H=1/2$ 只是其中的一个特例. 这类推广的随机无规函数被称为分形布朗运动. 通常用 $B_H(\Delta t)$ 来代替 $X(\Delta t)$ 作标记. 对于分形布朗运动，不难看出其均方值为

$$\langle \Delta B_H^2(\Delta t)\rangle \propto \Delta t^{2H}.$$

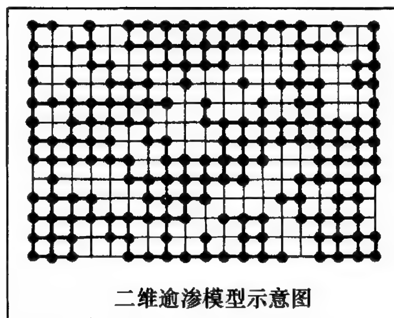
在时间标度作变化后，就可得到

$$\langle \Delta B_H^2(b\Delta t)\rangle \propto b\Delta t^{2H}=b^{2H}\langle \Delta B_H^2(\Delta t)\rangle.$$

也就是说，在时间轴上将 t 放大 b 倍后，函数轴上却是以 b^{2H} 的比率放大. 因此，分形布朗运动是一个自仿射分形. 利用分形布朗运动来研究自然、社会和经济现象中记录到的随机时间序列是有用的.

逾渗(percolation)　一种在随机几何结构中描述渗透过程的数学模型. 它是一种处理无序系统中各种物理问题与研究连续性相变的理论方法. 为了方便地描述流体在无序介质中的随机扩散与流动，1957 年，数学家布罗贝(Broadbent,S. R.)和哈默斯利(Hammersley,J. M.)提出了一种新的数学模型，其随机性来自于介质本身的无结构. 流体的这种行为类似于煮咖啡时通过渗滤壶(percolator)的流动，故哈默斯利称这种过程为逾渗过程.

以下列二维方形点阵为例来阐述逾渗模型. 设方形点阵上的格点可以随意地被占据，形如在围棋



二维逾渗模型示意图

盘上投放黑子，如图所示. 当黑子逐渐增多时，相邻的黑子就会构成大小不同的集团. 当达到某一个临界值时，点阵上一定会出现一个连通的集团；这个集团可以从点阵的底部扩展到顶部或者从左边扩展到右边，人们称这样的集团为无限大集团或逾渗集团. 当逾渗集团出现时就认为发生了一个逾渗相变. 它是一个简单的几何相变，反映的是通与不通这两种状态. 描述逾渗相变的各种参量都符合连续相变条件下的普适规律，所以逾渗相变是连续相变中的一例，连续相变理论为逾渗模型提供了有力的理论基础.

逾渗模型的主要应用有两类：一类是静态的逾渗相变，例如园中果树病虫害的防治、森林火灾的预防、地下沙岩的孔隙结构、材料裂纹的形成与断裂等；另一类是无序系统上动力学行为的研究，也就是在逾渗集团上的输运过程. 由于其结构的强无序就可带来反常的输运行为，通常可利用逾渗模型来研究各种无序材料中的反常扩散、反常电导等，最典型的是地下石油的开采、无规网络的电导计算等.

逾渗过程(percolation process)　见“逾渗”.

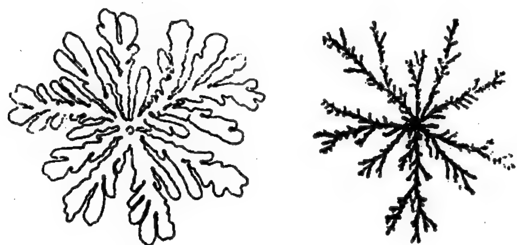
分形聚集(fractal aggregation)　一种基于分形的聚集生长模型. 若原有某个粒子可作为种子，于是其周围的粒子就不断地聚合成一个集团，由无规行走(扩散)方式聚合的粒子大多数聚合在集团的尖端附近，只有少量粒子进入到沟槽中，这样就显示出各种大大小小尺度的沟槽和触须，这种聚合产生的集团就称为分形生长或分形聚集.

该现象常见的例子如天空中的雪花. 它原有一冰晶作为种子，周围的水汽不断凝结、聚合而逐渐形成具有许多沟槽和六角形触须的分形雪花.

扩散置限凝聚模型(diffusion-limited aggregation model)　简称 DLA 模型. 分形聚集的一种计算机仿真模型. 1981 年，由美国科学家威顿(Witten,T. A.)和桑德尔(Sander,L. M.)提出. 它的生长规则是在一个二维方形点阵的中心处放置一粒种子，接着在离中心较远的边界处随机地产生一个粒子，并让它按无规行走方式作扩散运动. 当粒子运动到种子粒子的近邻处时就与粒子凝聚成一个集团，然后在边界处再产生一个粒子作扩散运动. 如此反复进行，直到生成如“分形”条目中图 3 所示的凝聚图形. 图形具有开放型的树枝状结构和无规自相似性质. 利用相关函数法求得其分形维数在二维与三维背景空间中分别为：$D\approx1.66\sim1.7$ 和 $D\approx2.5\sim2.53$. 自从 DLA 模型提出以后，引起了广泛的兴趣. 1984 年，许多科学家在电解沉积、气体放电、不同粘度流体运动的界面等不同领域的实验中都观察到 DLA 类的图形. 由于这些实验的支持，DLA 模型就成为一个重要的生长模型.

粘性指进(viscous fingering)　一种分形现象. 由两种粘滞性不同的流体相接触而产生. 当把一种粘性较大的流体(如甘油)放置于两块平板玻璃构成的容器中，并施加压力将另一种粘性较小的流体(如空气、水等)注入其间时，在容器内会出现一类树枝状的斑图. 这个实验由英国科学家海利-肖(Hele-Shaw,H. S.)在 1898 年首次完成，他的实验装置就称为海利-肖槽. 一般地，两种流体及其界面的运动遵守纳维-斯托克斯方程. 为处理这个极复杂的非线性方程，在忽略了若干作用项后，将其简化为拉普拉斯型方程. 从而在两种流体的界面处就出现了类似于扩散置限凝聚模型相似的分形斑图，这种现象称为粘性指进. 图示为二维平面上常见的粘性指进.

1984 年，皮特逊(Peterson，L.)证实粘性指进是一个分形，其分维数 D 约为 1.6 至 1.8.



典型的粘性指进图

粘性指进的研究对于石油开采具有很大的指导价值. 因为石油二次开采时须将水注入油田中进行驱油，这时若出现树枝状的粘性指进，将会使石油的采收率明显降低. 由此提出研究课题，如何让石油与水的界面上的表面张力尽可能小，以减小指进扩散，从而提高油井的采收率.

重整化群理论(renormalization group theory) 在粒子物理研究中为克服微扰发散困难而进行标度变换，从而得到群不变性的一种理论. 其后又被广泛用于研究凝聚态物理的相变问题. 重整化群理论可分为“动量空间重整化群”和“实空间重整化群”两大类. 重整化群的目的是通过改变物体的粗视化程度(长度标尺)来观察物体中各物理量的变化规律. 一个物体在发生二级相变的临界点处，它的相关长度是趋于无穷大的，因此物体就具有尺度变换下的不变性，也就是通常所说的“标度不变性”. 这时物体的结构必然具有自相似性. 可以利用标度不变性求出在临界点处的各种临界指数. 分形同样是一种具有自相似特性的几何体，它的结构满足标度不变性，因此基于标度不变特征的重整化群理论也是研究分形结构的一种有力工具.

重整化群理论由威尔逊(Wilson，K. G.)于 1971 年提出，并因而获得了 1982 年的诺贝尔物理奖. 重整化群计算方法一般可分为下列几个步骤：

1. 选择与划分基本元胞以便于进行标度变换.

2. 定义权重函数和求出重整化群变换公式，这是重整化群理论中的核心部分. 权重函数一般可选用多数法则或筛选法，其目的是减少体系的自由度. 重整化群变换公式是指同一物理量在不同尺度测量下的对应关系.

3. 确定重整化群变换的不动点，其依据是物体在临界点时其相关长度是趋于无穷的，因此相关长度在重整化变换下是一个不变量，即一个不动点.

4. 计算临界指数或分维数. 在不动点附近作线性近似后，求出各种临界指数的数值.

总之，重整化群理论是一个近似的理论，它只能对体系作粗视化处理而不能细视化. 尽管如此，它仍然是研究相变及许多非线性问题的一个有力工具.

自相似(self-similarity) 刻画无特征尺度形体的一种性质，即其整体与局部相似的一种性质. 整体与局部的相似性. 亦即当该集的任一个局部放大适当倍数后，它的形状将会和其原来的整体相一致. 其例包括寇赫曲线、谢尔品斯基地毯等(参见“分形”). 用数学语言可将自相似变换描述为：在欧氏空间中选一个点，设其坐标为

$$X = X(x_1, x_2, x_3, \cdots, x_N).$$

若将 X 各个分量的测量尺度改变 r 倍之后，其坐标变为

$$X' = X'(rX_1, rX_2, rX_3, \cdots, rX_N) = rX,$$

则称该变换为相似变换，其中 r 为实数，称为标度因子. 由于在各方向上的伸缩比相同，所以自相似变换具有伸缩不变性. 若所定义的变换沿各方向的伸缩率不完全相同，得到的变换是自仿射变换，即

$$X = X(x_1, x_2, x_3, \cdots, x_N),$$

但各方向上的伸缩比可能不同，它是一个比例向量，其表达式为

$$r = r(r_1, r_2, \cdots, r_N),$$

则自仿射变换可写为

$$X' = X'(r_1X_1, r_2X_2, r_3X_3, \cdots, r_NX_N).$$

一般地，自仿射变换是平移、旋转、伸缩和反射的合成，因此，它是自相似集的一种推广.

自仿射(self-affinity) 见“自相似”.

多重分形(multifractal) 一种分为多个区域的复杂分形结构. 为了对分形的复杂性和不均匀性进行更细致地刻画，需引进它的概率分布函数及其各阶矩的计算，由此构成了分形维数的一个连续谱，称之为多重分形或多标度分形. 由于多重分形至今尚无严格的数学定义，以下仅通过实例进行简单的描述. 设有一个分布不均匀的分形，首先将它分成若干个小区域，定义第 i 区域内的密度分布函数为 P_i，假定 P_i 存在着标度关系，即

$$P_i(\varepsilon) \triangleq \varepsilon^{\alpha}, \quad \alpha = \lim_{\varepsilon \to 0} \frac{\ln P_i(\varepsilon)}{\ln \varepsilon},$$

其中 ε 是一个小的测量单位，α 是一个标度指数，称为奇异指数，它是反映分形体内各个小区间的奇异程度的一个量，所以 α 的数值与其所在的位置有关. 进一步将分形上具有相同 α 值的小区间数记为 $N_\alpha(\varepsilon)$，则有

$$N_\alpha(\varepsilon) \triangleq \varepsilon^{-f(\alpha)},$$

$f(\alpha)$被称为奇异谱，因为它将奇异值 α 的密度用一个连续函数来表示. 另一方面，还可将分布函数进行加权求和，得到

$$\chi_q(\varepsilon) = \sum_{i=1}^{N} P_i^q = \varepsilon\tau(q),$$

$$\tau(q)=\lim_{\varepsilon\to 0}\frac{\ln\chi_q(\varepsilon)}{\ln\varepsilon},$$

$\tau(q)$常称为质量指数.再定义加权后的广义分形维数D_q为

$$D_q=\frac{1}{q-1}\lim_{\varepsilon\to 0}\sum\frac{\ln\sum_{i=1}^{N}P_i^q}{\ln\varepsilon}$$
$$=\frac{1}{q-1}\lim_{\varepsilon\to 0}\frac{\ln\chi_q(\varepsilon)}{\ln\varepsilon}.$$

从上述公式中可知,当$q=0$时,D_0就是普通的分形维数,而D_1称为信息维数,D_2称为关联维数.因此,D_q被称为广义分形维数.经过简单的计算可得α,$f(\alpha)$和$\tau(q)$,D_q这两组参量之间的对应关系式为

$$(q-1)D_q=\tau(q)=\alpha\cdot q-f(\alpha).$$

这一系列公式给出了多重分形理论的基本核心.利用多重分形,可把一个复杂的分形分成许多具有不同奇异程度的小区域来研究,从而能分层次地来了解分形的内部精细结构.

随机分形(random fractal) 采用随机生成机制而得到的分形集.分形体不具有特征尺度(亦即大小尺度跨好几个量级),却有自相似结构的分形.如康托尔集合,它由反复进行每次去其线段中间三分之一的操作而形成.用尺度r量出的积的数目就是

$$N(r)=r^{-D},$$

其中

$$D=\frac{\lg 2}{\lg 3}=0.6309,$$

就是康托尔集合的维数.

自然界中还有许多无特征尺度现象,它们虽不像康托尔集合那样是严格意义上的分形,但其统计特征和尺度在双对数坐标上仍可找到一个不变量.这时的分形就称为随机分形.例如,相距为r的两点的速度差$\Delta v=v(x+r)-v(x)$的统计特征,其二阶结构函数满足

$$\overline{[\Delta v(r)^2]}\triangleq r^{\frac{2}{3}},$$

这里数2/3是不变量,该式是湍流的一个重要特征.

朱利亚集(Julia set) 一类分形集.由法国数学家朱利亚(Julia,G.)于1918年提出.复平面上的简单迭代函数可写为

$$Z_{n+1}=Z_n^2+C\ \ (n=0,1,2,\cdots),$$

公式中的C是个给定的复数,Z是复变量.选定初值Z_0代入上式右端并反复迭代,即可得到序列$Z_0,Z_1,Z_2,\cdots$.其归宿有两种可能:

$$\lim_{n\to\infty}Z_n\to\infty,$$

或维持有限值.如果由Z_0出发的迭代趋于无穷,就在复数平面上把Z_0点涂白;否则将该Z_0点涂黑.这样每对应一个C值,人们便在复平面上得到一幅黑白图,黑白图的边界就是朱利亚集.下面给出几个朱利亚集的图形.

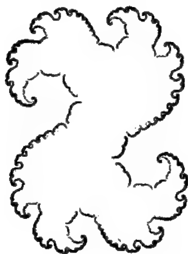


图 1

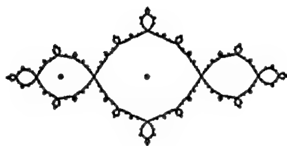


图 2

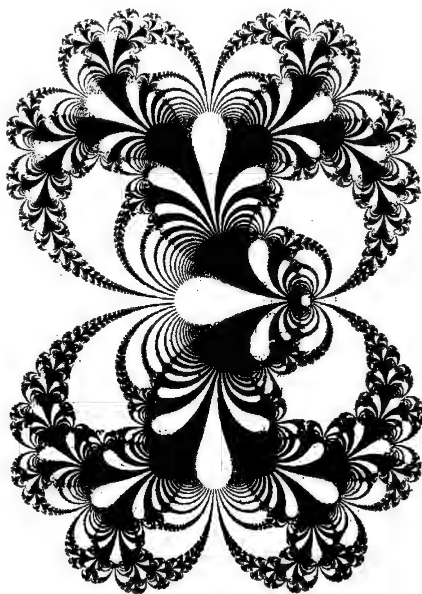


图 3

如果将上述图象的边界放大,则会出现更精细的结构,它们与原来的图象相似,这种局部中包含着整体的"全息"性质也是一种自相似性.一般地,朱利亚集的边界具有分形结构.

曼德勃罗特集(Mandelbrot set) 简称M集.由曼德勃罗特(Mandelbrot,B.B.)提出的一类分形集.曼德勃罗特集是在复平面上按

$$Z_{n+1}=Z_n^2+C\ \ (n=0,1,2,\cdots)$$

进行迭代,对应于一定的C值,在$n\to\infty$时,所有Z_n不趋于∞的Z_0点的边界构成了朱利亚集(参看"朱利亚集").从计算机模拟中还发现,参数C的微小变化都会造成朱利亚集形状的显著改变.在各种C值对应的朱利亚集中,有些是离散的,有些是连通的.如果在复平面上将对应于连通的朱利亚集的C点涂黑,而把对应于离散的朱利亚集的C点涂白,就会得到如下图所给出的点集.图中的黑色部分所代表的点集就是M集,边上的图形就是相应的M集中各种点C的朱利亚集.



M 集及其相应 C 点的朱利亚集

在复平面上将 C 值固定，而让 Z 作为原始点进行迭代变化，就可得到朱利亚集. 所以，对应于选定不同 C 值的朱利亚集有无穷多个，而 M 集只有一个. 如果把 M 集的边界放大，便会显示出 M 集下无穷尽的微型缩影. 这些微型 M 集与母集相似但并不完全一样. M 集的这种结构体现了分形几何的特点，即局部与整体的相似性. 这个性质与朱利亚集是一致的. 现在数学上已证明 M 集是连通的，即便看起来像是悬浮在平面上的 M 集也是通过细线同母集连结在一起的. 从某种意义上说，M 集概括了所有可能的朱利亚集，因此常把 M 集看成是朱利亚集的图象存贮器. 有关 M 集的性质仍是数学家们的一道难题.

费根鲍姆常数(Feigenbaum number)　非线性迭代映射的一个普适常数. 它反映了动力系统族中普遍存在的倍周期分岔过程中的某种内在的共同规律. 20 世纪 70 年代中期，费根鲍姆(Feigenbaum, M. J.)对一类单参数[0,1]区间上的一维映射迭代过程，其形式例如可为逻辑斯谛映射(参见“逻辑斯谛映射”)

$$x_{n+1} = f(x_n) \triangleq \lambda x_n(1 - x_n),$$

进行了深入研究. 他用计算机逐次计算映射出现倍周期分岔的 λ 值，发现分岔值序列按一种接近几何级数方式排列，如对上述例子，分岔值 λ_∞ 为常数，

$$\lambda_\infty = 3.569\,945\,672\cdots.$$

同时各 λ_i 值的间距比

$$\delta_n = \frac{\lambda_n - \lambda_{n-1}}{\lambda_{n+1} - \lambda_n},$$

当 $n\to\infty$ 时，δ_n 趋于极限值，

$$\delta = 4.669\,201\,661\cdots.$$

研究发现，对于一大类单峰映射该 δ 值均相同，从而对倍周期分岔具有普通意义，并被称为费根鲍姆常数.

逻辑斯谛映射(logistic map)　亦称抛物线映射. 表现复杂非线性行为的典型映射之一. 其具体形式为

$$x_{n+1} = \lambda x_n(1 - x_n) = f(x_n)$$
$$(0 \leqslant x_n \leqslant 1, 0 < \lambda \leqslant 4),$$

其中 x_n 代表第 n 步状态值. 由于函数 $f(x_n)$ 是抛物线(见图 1)，由图上可见 $x=1/2$ 时，$f(x_n)$ 达到最大值 $\lambda/4$. 抛物线映射迭代过程为：由任何初值 x_0 出发，按图上箭头进行，依次求出 $x_1, x_2, x_3, \cdots$，在图示例中就收敛到不动点 x^*，即抛物线和直线 $y=x$ 的交点. 当参数 λ 变化时，系统演化形态可以多种多样. 当 $0<\lambda<1$ 时，$x_n\to 0$，即 $x_n=0$ 是一个不动点. 当 $1<\lambda<3$ 时，$x_n\to 1-1/\lambda$，它将可能收敛到图上的另一个不动点 x_1^*. 这些均为过去熟知之结果. 但当参数 $\lambda>3$，可发现 $3<\lambda<3.45$ 时，x_n 在两个值上来回跳动，称为周期 2 解，它是映射 $x_{n+1}=f(f(x_n))$ 的不动点. λ 继续增大还有周期 4、周期 8 等. 直到 $\lambda_\infty=3.57\cdots$，系统状态为周期 2^∞，即非周期态，这就是由倍周期而导致的混沌状态(见图 2).

图 1

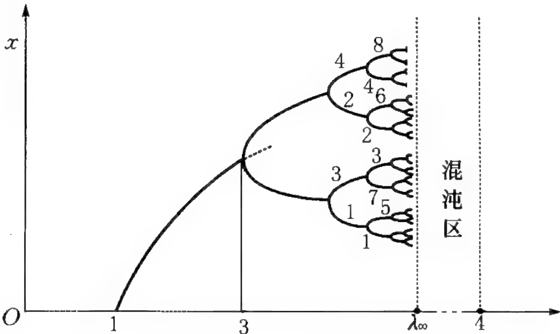


图 2

进而费根鲍姆(Feigenbaum, M. J.)还发现前面两个相邻分岔点之间的距离 $\lambda_n-\lambda_{n-1}$ 逐渐接近于后面两个相邻分岔点之间距离 $\lambda_{n+1}-\lambda_n$ 的 δ 倍，即

$$\delta = \lim_{n\to\infty} \frac{\lambda_n - \lambda_{n-1}}{\lambda_{n-1} - \lambda_{n-2}} = 4.669\,201\,661\cdots,$$

δ 称为费根鲍姆常数. 又发现周期 2^n 状态之间的距离 d_n 趋近于周期 2^{n+1} 状态之间距离 d_{n+1} 的 α 倍，即

$$\lim_{n\to\infty} \frac{d_n}{d_{n+1}} = \alpha = 2.5\cdots,$$

α 也是无理数.

抛物线映射(parabola map)　即“逻辑斯谛映射”.

帐篷映射(tent map)　一种分段线性映射. 其具体形式为

$$x_{n+1} = f(x_n) = \begin{cases} \lambda(1 - x_m)x_n & (x \leqslant x_m), \\ \lambda x_m(1 - x_n) & (x_m \leqslant x \leqslant 1), \end{cases}$$

见图. 因其形似帐篷而得名. 取 $\lambda=4$，就可使其在极大点 $x_m=1/2$ 处取值为 1 的对称的帐篷映射.

S 映射(S-map)　一种简单的一维非线性映射. 其具体形式为

$$x_{n+1}=f(x_n),$$

它由间隔[0,1]到其自身的映射中，仅在 x_m 处有一个最大值，并且 $f(x)$ 处处有负的施瓦兹导数，即

$$\frac{f'''}{f'}-\frac{3}{2}\left(\frac{f''}{f'}\right)^2<0.$$

显然，逻辑斯谛映射 $x_{n+1}=\lambda x_n(1-x_n)$ 也是一种 S 映射. 辛格(Singer, D.)证明，S 映射对每个参数 λ 至多存在一个稳定的周期解，且若它存在，可以从 x_m 迭代求出，即 x_m 在周期解的吸引盆上.

圆映射(circle map)　一种能展现多种混沌特征的一维映射. 其数学形式为

$$x_{n+1}=x_n+\Omega+f(x_n),$$

其中 Ω 是整数，$f(x)=f(x+2\pi)$ 是周期函数，它是映射的非线性部分，Ω 是非线性项不存在时映射的频率. 在周期驱动下的有阻尼单摆强迫振动问题中，常化为在圆(0 到 2π)上的圆映射. 在圆映射上，若 $f(x)$ 是正弦函数时取如下形式：

$$\theta_{n+1}=\theta_n+\Omega-\frac{k}{2\pi}\sin(2\pi\theta_n),$$

则称其为标准映射，其中 k 为耦合参数.

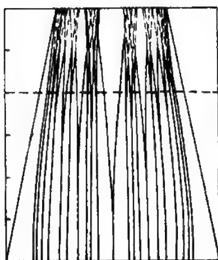


图 1



图 2

在驱动阻尼单摆系统中有两个特征频率，一个是单摆的自然频率，另一个是驱动频率. 当两者耦合比较弱时(例如 $k=0$)，则映射成为线性的，$\theta_{n+1}=\theta_n+\Omega$. 这时从任何初条件 θ_0 出发，或形成周期解(Ω 为有理数时)，即在圆上绕若干圈后回到原处；或为概周期解(Ω 是无理数时). 当两者耦合比较强时(k 较大，例如 $k=0.95$ 接近于 1)，也可以发生锁频现象，即映射迭代 q 次以后，角度和初始角相差一个整数 $\theta_{n+q}=\theta_n+p$，这里 p,q 是正整数，p/q 是有理数，作为锁频现象的特征，引进旋转数 ω,

$$\omega=\lim_{n\to\infty}\frac{\theta_n-\theta_0}{n},$$

它代表每单位时间的平均角度增长(平均频率). 锁频时，$\omega=p/q$ 为有理数；而当 ω 为无理数时，代表准周期状态；ω 不确定时为混沌状态. 上述映射频率锁相的锁定范围的宽度与耦合参数 k 有关，图 1 阴影区为锁定区域(在 k-Ω 图上)，它们用苏联数学家阿诺尔德命名，称为阿诺尔德舌头. 在 $k=1$ 的临界值时，阿诺尔德舌头的宽度增大到开始彼此衔接. 此时旋转数 ω 与 Ω 的关系是由无数个大大小小的平台和阶梯组成，称为魔鬼阶梯. 图 2 为魔鬼阶梯及与其对应的康托尔集.

标准映射(canonical map)　见"圆映射".

阿诺尔德舌头(Arnold tongue)　见"圆映射".

魔鬼阶梯(devil staircase)　见"圆映射".

厄农映射(Henon map)　可以产生混沌现象的一种二维映射. 由法国天文学家厄农(Henon, M.)于 1976 年提出，其表达式为

$$x_{n+1}=y_n+1-ax_n^2,$$
$$y_{n+1}=bx_n,$$

它的雅可比行列式

$$|J|=\begin{vmatrix}\frac{\partial x_{n+1}}{\partial x_n} & \frac{\partial x_{n+1}}{\partial y_n}\\ \frac{\partial y_{n+1}}{\partial x_n} & \frac{\partial y_{n+1}}{\partial y_n}\end{vmatrix}=\begin{vmatrix}-2ax_n & 1\\ b & 0\end{vmatrix}=-b,$$

若 $|b|<1$，该映射代表面积收缩的映射(即耗散系统)，称为厄农映射. 在计算中通常取 $a=1.4$，$b=0.3$. 若 $b=-1$，则是厄农面积守恒的二维映射. 1979 年，他又提出另一个面积守恒的二维映射：

$$x_{n+1}=(\cos\alpha)x_n-\sin\alpha(y_n-x_n^2),$$
$$y_{n+1}=(\sin\alpha)x_n+\cos\alpha(y_n-x_n^2).$$

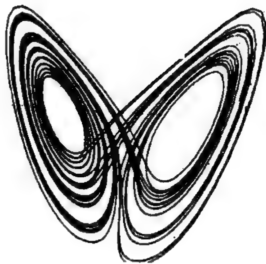
它们对研究保守系统的混沌及计算方法均有很大价值.

洛伦兹方程(Lorenz equation)　描述空气流体运动的一个简化微分方程组. 1963 年，美国气象学家洛伦兹(Lorenz, E. N.)将描述大气热对流的非线性偏微分方程组通过傅里叶展开，大胆地截断而导出描述垂直速度、上下温差的展开系数 $x(t)$，$y(t)$，$z(t)$ 的三维自治动力系统：

$$\frac{\mathrm{d}x}{\mathrm{d}t}=-Prx+Pry;$$
$$\frac{\mathrm{d}y}{\mathrm{d}t}=-Rax-y-xz;$$
$$\frac{\mathrm{d}z}{\mathrm{d}t}=xy-\frac{8}{3}z.$$

其中，Pr 为普朗特数，Ra 为瑞利数. 他发现当 Ra 不断增加时，系统就由定常态(表示空气静止)分岔出周期态(表示对流状态)，最后，当 $Ra>24.74$ 时，又分岔出非周期的混沌态(表示湍流). 如图是三维相空间的混沌态在二维平面上的投影轨线. 从图上看出，轨线起初在右边从外向内绕圈子，后来随机地跳到左边从外向内绕圈子，后又再次随机地跳回右边绕圈子，……如此左右跳来跳去，每次绕的圈数，

何时发生跳跃都是随机的、无规则的.



由于洛伦兹是世界上第一个从确定的方程中发现了非周期的混沌现象，所以将上述方程一般称为洛伦兹方程.

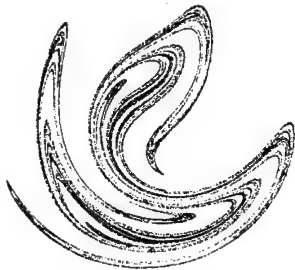
蝴蝶效应(butterfly effect)　混沌系统对初值敏感依赖的形象比喻.美国气象学家洛伦兹(Lorenz,E.N.)从多年气象研究中认识到长期天气预报的不准确性.他由几乎相同的初始条件出发，用同一气象模型(洛伦兹方程)，却预报出完全不同的结果.他形象地将初始条件的微小差别比作蝴蝶的翅膀一拍，其结果将对天气预报造成很大误差.所谓蝴蝶效应，就是将上述误差比喻成北京的一只蝴蝶翅膀一拍，就会影响纽约天气的长期变化.究其原因就在于确定性非线性系统中混沌现象的内在不稳定性，即对初条件高度敏感的特性：初始条件稍有差别，运动轨道的差别就很大.

间歇混沌(intermittent chaos)　系统通向混沌的一种方式.它指系统形成混沌过程中，呈现出的有时为混沌、有时为周期的状态，亦即在混沌的轨道中可能要经过看起来不像混沌的阶段，这是通向混沌的道路之一.例如，逻辑斯谛模型

$$x_{n+1}=\mu x_n(1-x_n)$$

中，当 $\mu=1+\sqrt{8}=3.828$ 时，会出现周期3.但当 μ 接近、但稍稍小于 $1+\sqrt{8}$ 时，其轨道中会出现一段好像是周期3的周期状态，然后又进入无周期的混沌状态，从而其轨道就貌似周期和混沌的间歇发生.

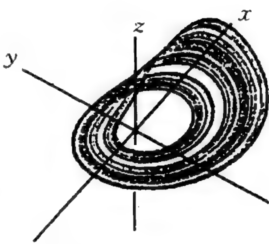
日本吸引子(Japan attractor)　亦称上田吸引子.一种奇怪吸引子.它是由日本学者最先发现的.现在，世界上公认首先发现奇怪吸引子的是美国气象学家洛伦兹(Lorenz, E. N.)在1963年发现的洛伦兹吸引子.但实际上，京都大学上田皖亮早在1961年就由周期强迫的杜芬方程发现了奇怪行为的证据.但由于没有高性能的计算机，他未得到更高分辨的图象和更深入的结果.后来法国数学家茹勒(Ruelle,D.)把上田皖亮的吸引子称为日本吸引子.该吸引子如图所示.



吕兹勒吸引子(Rössler attractor)　一种奇怪吸引子.它是由德国科学家吕兹勒(Rössler,O.)提出的.他曾用下列非线性方程组

$$\dot{x}=-y-z,$$
$$\dot{y}=x+\alpha y,$$
$$\dot{z}=\alpha+xz-\mu z$$

作为化学反应的模型.当取 $\alpha=0.2,\mu=5.7$ 时也产生奇怪吸引子.注意该方程组中只有一项 xz 是非线性项，比洛伦兹方程中还要简单.吕兹勒吸引子如下图所示.



康托尔集(Cantor set)　具有分数维数的一种简单集合.它早已由德国数学家康托尔(Cantor, G.)于1883年提出.其构造法则如下：由单位长度区间[0,1]记为 A_0 出发，去掉其中段1/3，得到的集记为 A_1，它包含两个子区间[0,1/3]和[2/3,1]，再接着去掉 A_1 的两个子区间各自中段1/3得到集 A_2，按此法则操作 k 次后得到集 A_k，它由 2^k 个长度为 3^{-k} 的小区间组成，记为

$$A_0=[0,1],$$

$$A_1=[0,\frac{1}{3}]\cup[\frac{2}{3},1],$$

$$A_2=[0,\frac{1}{9}]\cup[\frac{2}{9},\frac{1}{3}]\cup[\frac{2}{3},\frac{7}{9}]\cup[\frac{8}{9},1],$$

……………………………………………

因此，康托尔集可以写成

$$A=\bigcap_{k=0}^{\infty}A_k=\lim_{k\to\infty}A_k.$$

康托尔集 A 是不可数的，因为它是一个非空的完备集.若计算一下从闭区间[0,1]中去掉的点集的勒贝格测度(即各区间的长度)，得到

$$\frac{1}{3}+\frac{1}{3}\cdot\frac{2}{3}+\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}+\cdots$$

$$=\frac{\frac{1}{3}}{1-\frac{2}{3}}=1,$$

所以康托尔集 A 的勒贝格测度等于零.其分维数则为0.63….

$1/f$ 噪声($1/f$ noise)　系统中出现以振动频率 f 某一幂次表征的发散现象.在早期的无线电二极管和晶体管实验中，可观察到其电流的功率谱 S 在低频处随振动频率 f 以幂次规律发散现象，后来又在许多系统中，如半导体和水晶片的振动、高速公路上车辆流动、气温的季节变化、音乐、人体的心电图以及经济领域中的股票价格的变动等，观察到同样的幂次规律发散现象.科学家把这类现象称为 $1/f$ 噪声，其功率谱 S 可表达为

$$S=\frac{1}{f^{\beta}}\quad(0.8<\beta<1.4).$$

有时也把这类噪声称为闪变噪声或粉红色噪声. 这种说法也是相对于人们熟悉的白色噪声($\beta=0$)和高斯噪声($\beta=2$)而言的. 尽管在如此多的系统中都可以观察到 $1/f$ 噪声. 对于产生 $1/f$ 噪声的原因，现已达到的共识为：

1. 凡具有 $1/f$ 噪声的系统，其概率分布都服从稳定的统计分布，即对数正态分布.

2. $1/f$ 噪声的波谱具有自相似的特性，在时间尺度上满足标度不变性.

3. 当一个系统的时间相关长度或空间相关长度趋于无限大时，该系统就会出现 $1/f$ 长时尾效应.

但以上各点只是科学家们在认识 $1/f$ 噪声基本机理过程中的归纳. 由于 $1/f$ 噪声对自然界中的噪声和生物系统的研究具有重要意义，人们仍在继续探索.

闪变噪声(flicker noise) 见“$1/f$ 噪声”.

沙可夫斯基定理(Sharkovsky theorem) 关于一维映射 $x_{n+1}=f(x_n)$ 的周期解性质的命题. 1964年，沙可夫斯基(Sharkovsky，A. N.)证明，对闭区间[0,1]映射到本身的映射，存在下面实数序列：

$$1,2,2^2,2^3,2^4,\cdots;$$
$$\cdots 7\times 2^3,5\times 2^3,3\times 2^3,\cdots;$$
$$\cdots 7\times 2^2,5\times 2^2,3\times 2^2,\cdots;$$
$$\cdots 7\times 2,5\times 2,3\times 2,\cdots;$$
$$\cdots 7,5,3.$$

若映射有周期为某数的解，那么就一定有上述序列中排在它前面的那些数的周期解. 例如，对于映射 $x_{n+1}=\mu x_n(1-x_n)$，若有周期为 16 的解，那么一定有周期 8、周期 4、周期 2、周期 1 的解. 同样，若有周期 3 的解，由于 3 是上述序列的最后一个数，因此映射就有它前面的所有周期解. 1975 年，李天岩和约克(Yorke，J. A.)发表文章，题目是“周期 3 意味着混沌”就是这个定理的含义.

塔肯斯定理(Takens theorem) 计算嵌入相空间维数大小的一个基本命题. 它是重构相空间技术的理论依据. 混沌应用的一个重要问题就是从单个变量的时间序列重新构造一个可包容该混沌运动的 n 维相空间. 对于维数 n 究竟应该取多大的问题，1934 年，美国数学家惠特尼(Whitney，H.)；1980年，塔肯斯(Takens，F.)先后证明了所需维数 n 大小的嵌入定理：为了保证该相空间容纳该状态空间原来吸引子的拓扑特征，如果原来吸引子处在一个 d 维空间中，那么，将该吸引子嵌入其中的相空间维数必须达到 $n=2d+1$. 例如由某时间序列重构相空间所求得的吸引子的维数为 $D=2.51$，亦即该吸引子处在 $d=3$ 维的空间中，因此嵌入相空间的维数至少为 $n=2\times 3+1=7$.

不变测度(invariant measure) 测度的一种. 对于出现混沌的一维映射

$$x_{n+1}=f(x_n) \quad (x_n\in[0,1]),$$

虽然混沌运动轨道的点集 $\{x_n\}$ 具有对初条件敏感的性质，但是它可以用一个分布函数 $P(x)$ 来描述. 即在区间[0,1]上的每一 x 处都有一定的概率测度

$$\mu \mathrm{d}x = p(x)\mathrm{d}x,$$

该测度 $p(x)$ 随着由 x_n 到 x_{n+1} 的迭代映射，亦相应地演化. 显然这时存在一个和映射 f 有关的算子 $\mathscr{F}$ (称 Frobenias-Perron 算子)，使该测度由 μ 演化到 $\mathscr{F}\mu$，再变到 $\mathscr{F}^2\mu$ 等. 这时方程 $\mu=\mathscr{F}\mu$ 的不动点就是不变测度. 由于 x_n 是 x_{n+1} 的逆映射，因此不变测度也意味着

$$\mu(x)=\mu(f^{-1}(x)).$$

因为存在不变测度，可知混沌轨道是各态历经的，此时时间平均就可以用按概率分布的平均来代替.

决策与博弈系统

博弈论(game theory) 亦称对策论. 研究人类竞争和合作行为的数学理论. 在两个以上决策人共同做决策的环境中，各人力求使自己的效用函数达到极大的决策行为，其数学描述包括以下几个方面：

1. 设有 $N(\geqslant 2)$ 个决策人或局中人 $\mathrm{DM}_1,\mathrm{DM}_2,\cdots,\mathrm{DM}_N$. 第 i 人 DM_i 的决策量为 u_i，其取值范围为 U_i，$u_i\in U_i$，U_i 可以是有限集、某区间或函数空间(这时 u_i 可为某时间函数，对应于微分对策问题).

2. 每个 DM_i 的效用函数(常称目标函数) $J_i(u_1,u_2,\cdots,u_N,\xi)$ 依赖于其他各人的决策，还可能有随机变量 ξ.

3. 每个 DM_i 做决策时可以依据的信息 $y_i=\eta_i(u_1,u_2,\cdots,u_N,\xi)$，它可能是带随机干扰的对系统状态或他人行动的观测.

4. 各 DM_i 由信息 y_i 确定其决策量的规则，即策略 $\gamma_i\in\Gamma_i$，亦即 $u_i=\gamma_i(y_i)$，而 Γ_i 表示可取函数类，例如要满足因果性条件或其他数学条件(约束)等. 特别应当指出，Γ_i 还可能包括按某概率分布随机选取 u_i 的情形(参见“混策略”).

5. 规定对策规则，如各人决策的逻辑或时间顺序等. 对策的描述总的可分为两种形式：正规形式是要求对策人一次性地按公布的规则选定自己的全部策略，并同时摊牌以决定结局；展开形式则按逻辑次序指明对策各方轮番决策过程的树状结构，各决策人应依次确定自己的决策，逐步得出结局. 显然，后一形式更加完备，可以导出前者，但前者也具有简单明了的优点.

6. 理性准则，即关于各人行为特征的说明，如冒险或保守，倾向于合作或对抗等.

按上述描述，对策问题可以划分为许多子类，并具有各自不同的解概念. 其主要类别有两人或多人、零和(亦称常和，指 $J_1+J_2+\cdots+J_N=$ 常数，表示完全对抗情形)或非零和、静态(一次性)或动态(u_i 是时间函数，决策人之间有相互响应的问题)等. 对策论按其研究角度又可分为两类. 其一是规范理论，从某些公认或假定的行为规范出发，推导出决策人的行为应当怎样；另一则为描述理论，即从观察到的行为研究它们所依据的理性准则. 对策论的内容除前述完全对抗的两人零和对策外，非零和对策中又可分为非合作平衡解、半合作的帕雷托优化和协商理论和多人合作的结盟理论，以及当决策人分属不同层次时的主从对策和激励对策等. 伴随时间演化的微分对策(亦称动态对策)也相应地具有上述各种类型.

对策论已被广泛用于社会、经济、军事、企业管理等各种人类活动系统，并日益得到重视. 它的概念和方法还可用于研究多智能体模型和各种人类征服自然的活动. 对策论虽然早在 1921 年曾由波莱尔(Borel,(F.-É.-J.-)É.)提出，它的主要结果和理论体系则是由美籍匈牙利数学家冯・诺伊曼(von Neumann,J.)和莫根施特恩(Morgensten,O.)奠定的. 微分对策则更晚些，由艾萨克斯(Issacs,R.)于 1965 年提出.

效用(utility)　决策结果有用程度的一种定量度量. 对于相同的损益期望值，不同决策人的反应不一定相同. 最后做出的决策是由决策人的个人性格、当前处境、对未来的展望等条件形成的. 即使是同一个决策人，由于时期、条件等不同，对相同的利益或损失的反应也不一定相同. 决策人对于利益和损失的独特的兴趣、感觉或反应综合起来称为效用. 效用实质上也代表了决策人在不确定的决策环境中对于风险的态度.

极小-极大后悔准则(min-max regret criterion)　不确定型决策准则之一. 它要求决策最大后悔值尽可能地小. 这里，最大后悔值指不确定条件下某决策可能得到的最坏结果与其他决策所可能得到的最好结果之差.

极大-极大准则(max-max criterion)　不确定型决策准则之一. 它倾向于乐观的结果，并以各个方案可能导致的最好结果作为决策的依据. 设方案 a_i 导致结果 c_j 时其所得为 P_{ij}，则极大-极大准则要求选择 a_{i^*}，使达到

$$\max_i \max_j P_{ij}.$$

极大-极小准则(max-min criterion)　不确定型决策准则之一. 它倾向于保守，以各个方案可能导致的最坏结果作为决策的依据. 设方案 a_i 导致结果 c_j 时其所得为 P_{ij}，则极大-极小准则要求选择 a_{i^*}，使达到

$$\max_i \min_j P_{ij}.$$

拉普拉斯准则(Laplace criterion)　亦称等可能性准则或平均准则. 不确定型决策准则之一. 它处理不确定型决策问题的思路是每种方案对应多种结局，在不知道各种结局 c_i $(i=1,2,\cdots,m)$ 和发生概率的情况下，认为各种结局出现的概率 $p(c_i)$ 等同，相当于认为

$$p(c_1)=p(c_2)=\cdots=p(c_m)=\frac{1}{m}.$$

因此，各方案 a_j 的期望效益值 $V(a_j)$ 为

$$V(a_j)=\frac{1}{m}\sum_{i=1}^{m}v(c_i,a_j),$$

其中，$v(c_i,a_j)$ 为方案 a_j、结局 c_i 情况下的效益值. 这种决策准则下的最优方案 a_{j^*} 应该满足

$$V(a_{j^*})=\max_{a_j\in A}V(a_j).$$

理想点(ideal point)　一种最优状态. 指理论上存在但实际无法达到的状态. 在多目标优化问题中，由于目标间存在矛盾性，各目标难以同时达到最优值. 在数学上，对理想点的表述是每一个目标函数在目标空间上存在一个解向量，把每列最大值 f_i^* $(i\in n)$ 所构成的向量

$$f^*=(f_1^*,f_2^*,\cdots,f_n^*)$$

称为理想点. 一般地，f^* 可能并不处于可行域内. 这时求解多目标优化的方法之一就是在可行域中找出一点，使其在某种意义下与理想点最为接近.

满意解(satisfactory solution)　一种非最优解. 在处理实际系统优化过程中，由于人们对系统结构、状态、参数了解不充分，或对于系统信息掌握不完备，要求得系统严格的最优解是不现实或不必要的. 一般地，在实际工作中只要在可行解集合中找到一个决策者满意的解就可以了，这种解就泛称为满意解.

有限理性原理(principle of bounded rationality)　寻找满意解的理论依据. 由西蒙(Simon,H.)提出，说明人类决策不可能简单地归结为某种目标函数优化的完美数学形式的原理. 它的原因在于人类目标的模糊性，其知识和信息的不完备性，其推理判断能力的局限性等. 西蒙由于这一理论及其成功应用获得了 1978 年的诺贝尔经济学奖.

多目标决策(multi-objective decision-making)　亦称多目标优化. 考虑多种目标同时实现优化的决策方法. 在复杂大系统中，有可能由具有不同目标的多人共同决策，有待优化的目标较多，而且相互牵制. 多目标决策的问题是较普遍的. 处理多目标决策的方法很多. 首先是求非劣解即帕雷托解集，即任一

目标均不能在不损害其他目标的条件下进一步改善的那些解的集合，然后再用附加的理性准则由该集合中确定一个解答. 在各目标具有不可比的重要程度时，可以用层次分析法等确定其优先级，先对高层、重要的目标进行优化，然后再逐级优化其他目标. 若各目标之间可以比较，则可以用加权求和等方法将其化为一个单目标优化问题来求解.

多目标优化(multi-objective optimization) 即“多目标决策”.

帕雷托解(Pareto solution) 亦称非劣解或有效解. 多目标优化的一种解概念. 由意大利经济学家帕雷托(Pareto, L.)提出. 设有 N 个决策人，第 i 人的决策量为 u_i，其极大化目标函数为 $J_i(u_1, u_2, \cdots, u_N)$. 称点 $u^p=(u_1^p, u_2^p, \cdots, u_N^p)$为一帕雷托解，是指不可能找到更好的决策 $u^*=(u_1^*, u_2^*, \cdots, u_N^*)$，使得对任何 i，均有 $J_i(u^*)\geqslant J_i(u^p)$，且其中至少有某 i 使得不等号严格成立. 由于帕雷托解一般并不惟一，如何由其中按某种理性准则选出最满意的结果，将是这一类多目标优化问题求解的关键.

非劣解(non-inferior solution) 即“帕雷托解”.

多人决策系统(multi-person decision-making system) 决策系统概念在多决策人相互影响的情形下的推广. 由于各决策人可有不同的目标函数，掌握不同的信息和决策权限，依照决策规则而居于不同的地位，多人决策系统的问题比单人决策情形要复杂得多. 多人决策问题的详细提法及其主要内容和研究方法，可参见“对策论”. 但其中有一种特殊情形，即各人目标函数均相同时，又称为组队决策问题或群决策问题(参见“组队决策”). 它的解与信息结构，即各决策人相互通信并拥有信息的情况有密切关系.

效用理论(utility theory) 研究人类决策时作为极大化目标的效用函数的理论. 通常认为，财物收益、精神满足等总可折算为某种数量化的指标. 在确定性情形下，任何有一一对应换算关系的不同指标所导致的决策结果亦应相同. 但若存在风险因素，情况则完全不同. 常见的例子如，抽签以 0.5 概率获得一大笔钱，不如以概率 1 稳拿其中半数. 由此可知，钱数和效用值并非正比. 由此形成的效用函数概念最早由克拉美(Cramer, G.)和丹尼尔·伯努利(Bernoulli, Daniel)于 1738 年提出，但直到 1944 年才由美籍匈牙利数学家冯·诺伊曼(von Neumann, J.)和莫根施特恩(Morgensten, O.)建立了严格的公理化系统. 如果把所有可能结果的概率分布称为“情景”P，则效用函数可定义为 P 上满足以下公理的实值函数 U：

1. U 的大小符合该决策问题的优劣序.

2. U 在 P 上具有线性.

费史伯恩(Fishburn, P. C.)于 1972 年、詹森(Jensen)于 1967 年先后给出了优劣序应该满足的公理，以保证效用函数的存在. 效用函数的形状反映着决策人对风险的喜恶倾向，例如，凸函数表示厌恶风险的保守倾向，凹函数表示冒险倾向等.

还可以定义多因素的效用函数，例如付出多少劳动得到多少报酬的满意程度，或者购物时对商品多种属性的综合评估等. 总之，这是人类决策必须明确的首要问题.

多人决策系统理论(multi-person decision-making system theory) 具有多决策人的决策系统和决策过程的理论. 在多人相互影响条件下，群体决策和博弈系统的主要特征可概括为以下方面因素：

1. 设共有 N 个决策人，分别记为 DM_i ($i=1, 2, \cdots, N$)，其决策分别为 $u\in U$，在动态情形可能是个时间函数.

2. 各决策人追求的目标是使自己的效用函数 J 达到最大，但 J 也受其他人决策和随机因素的影响.

3. 各 DM 决策时可以依据的信息 $y=\eta(u_1, u_2, \cdots, u_N, \xi)$，这是依靠观测和通信获得的，可能还有相当的代价须计入成本.

4. 各 DM 的策略 $u=\gamma(y)$，是对信息 y 进行加工以确定决策量的规则，γ 应满足一些数学可操作性的条件，以便于比较、判断.

5. 规定游戏规则，包括各人决策行动的逻辑或时间顺序，以及决策的方式和内容，例如，宣布其策略，提供某些信息，采取某行动及其顺序规则等.

6. 此外，通常还希望对决策人的行为特征做些尽量明确的说明，通常包括冒险或保守、合作与对抗等因素，它们可能根本改变对策的结果.

由此可定义多人决策或对策的各种较明确的“解概念”，诸如组队决策、两人零和对策、非零和对策、动态对策、多人(结盟)对策、主从激励对策等.

博弈模拟(gaming) 用计算机模拟和人类交互来研究博弈的一类方法. 有两个或两个以上局中人(其中一方可能用设计好的计算机软件来代表)参与的博弈的一种实现，而且这种实现可以用博弈模型来描述和评估. 根据博弈模拟的目标的不同，可将模拟分成：

1. 教学模拟. 这是大多数博弈模拟的主要目的. 在培养作战指挥员、工商管理人员或专门人材的学校里，博弈模拟常作为训练学员了解基本的操作原理并教会他们应变和对抗“技能”的手段.

2. 操作模拟. 其目标是模拟较复杂的实际决策问题，演示可能的结果，以求找出正确的解答. 在操作模拟中，局中人通常就是实际决策者，模拟的设计

应使模拟过程尽可能接近实际情况.

3. 研究模拟. 用来检验在冲突或合作局势下，人们的决策实际上是否符合专家或心理学家对局势所做的分析. 这种模拟力图尽可能地接近一个可重复的试验，其目的是收集数据，分析应试者的行为，以检验一个特定的概念或假设.

4. 娱乐模拟. 一种最古老的博弈模拟. 它包括了人们最流行的竞赛性娱乐活动，如各种球类、棋类活动及赌博游戏. 近年来，娱乐模拟的最大发展是电子计算机游戏. 这种博弈模拟看起来只有一个局中人，但设计者和随机数发生器实际上代表了另外各方. 设计这种游戏正是博弈论可以充分发挥作用的领域.

一个博弈模拟试验的过程一般分为模拟设计、模拟准备、模拟实施和模拟后分析等四个阶段.

组队决策(team decision)　一类多人决策问题. 其中各决策人的目标函数是一致的，均为 $J(u_1,u_2,\cdots,u_N)$. 尽管他们为了共同的目标可以事先商定行动计划或控制法则，但由于存在不确定因素，他们在行动时掌握的信息可能不同，从而难以协同动作而达到最佳结果. 尽管这可以由及时相互通信交换信息来解决，但又带来了费用、时延、噪声等不利影响. 究竟每个决策人行动时掌握多少信息，即信息结构问题，就成为组队决策特有的关键问题. 组队决策问题中决策人 DM_i 的信息 y_i，无非是与随机噪声 ξ 和其他人的决策量 $u_j(j\neq i)$ 有关的量. 若 y_i 与其他人的 u_j 均无关，则称之为静态信息结构. 若 y_i 依赖于其他人的 u_j，则意味着 DM_i 要在 DM_j 之后行动. 这时，如果先行动者 DM_j 的信息 y_j 并不依赖于后行动者 DM_i 的作用 u_i，称这种信息结构为部分嵌套的，它并不违背两者间的因果关系. 最一般情形，即某些 DM_i 和 DM_j 的信息与对方的决策量有相互依赖关系，则称为非嵌套的信息结构. 无论是部分嵌套和非嵌套的信息结构，均称为动态的. 而所有各 y_i 不完全相同的信息结构，均称为非古典的. 组队决策理论的基本结果是：对静态或部分嵌套信息结构，线性二次高斯的组队决策问题是可解的，而且各决策规则 $u_i=\gamma_i(y_i)$ 均为线性函数. 而对于非嵌套的动态信息结构尚无一般解法，威腾豪森(Witsenhausen, S.)曾在 1968 年对此给出过一个著名的反例. 这类问题与分散随机最优控制亦有密切关系. 组队决策概念最早由经济问题中提出，后来又广泛应用于各种分散信息处理和控制决策系统研究之中，其典型例子有分布式数据库系统设计、指挥-控制-通信-情报(C^3I)系统等.

信息结构(information structure)　多人决策中有关信息的分布结构. 多人动态决策和对策问题中，除了目标函数、游戏规则外，各人决策时可依据的信息，亦即信息结构，是个至关重要的问题. 在大家目标完全一致的组队决策问题中尤其突出，整个优化问题实际上是系统品质与信息传输、处理的费用及安全、保密方面的代价之间的一种折衷. 从大系统分散随机控制理论的角度看来，不良的信息结构可能导致品质低下，甚至整个问题的不可解性.

信息嵌套原理(information nest principle)　一种关于组队决策信息结构的原理. 作为组队决策问题可解性的一个条件，信息嵌套原理要求，若决策人 i 的决策受到决策人 j 的影响，则 i 必须了解 j 所拥有的信息. 亦即 j 的信息还须被嵌套在 i 的信息之中，这时“非古典信息结构的线性-二次型-高斯随机优化问题”才是可解的.

协商理论(bargaining theory)　亦称谈判理论. 研究合作对策的理论之一. 协商各方愿意恪守诺言，从而可实现某种合作以达到较好的结果. 协商理论的内容包括各种理性准则及由其导出的协商条件和程序. 以纳什(Nash, J. F.)的协商理论为例. 设两个所有可能的决策对 (u_1,u_2) 导致的支付函数 $(J_1(u_1,u_2),\ J_2(u_1,u_2))$ 构成平面上某集合 Ω. 协商的目的是在可行集 Ω 中选定一个双方认为合理的点，并用相应的 (u_1,u_2) 加以实现. 所谓“合理”，纳什建议用以下六条公理来描述：

1. 可行性，$(J_1^*,J_2^*)\in\Omega$.

2. 个人合理性，即各人所得不劣于完全不合作的结果，即 $(J_1^*,J_2^*)\geqslant(\bar{J}_1,\bar{J}_2)$，其中

$$\bar{J}_i=\max_{u_i}\min_{u_j}J_i\ (i,j=1,2;i\neq j).$$

3. 非劣性(Pareto 最优性)，不应当能找到双方均可改进的结果，亦即若 $(J_1,J_2)\in\Omega$，$(J_1,J_2)\geqslant(J_1^*,J_2^*)$，则必有 $J_1=J_1^*$，$J_2=J_2^*$.

4. 无关于多余的抉择. 若 $(J_1^*,J_2^*)\in\Sigma\subset\Omega$，并设公理 2 的要求在 Σ 范围内亦满足，则以 Σ 代替 Ω 时协商解应当不变.

5. 对 J_1 或 J_2 分别做线性变换，不改变协商的结果.

6. 对称性. 若 Ω 是对称集，即若 $(J_1,J_2)\in\Omega$，则 $(J_2,J_1)\in\Omega$，则协商之结果有 $J_1^*=J_2^*$.

在上述公理下，纳什证明了他关于协商的基本定理：对确定的可行集和公理 2 确定的下界 $(\bar{J}_1,\bar{J}_2)$，存在惟一的协商结果 $\varphi(\Omega,\bar{J}_1,\bar{J}_2)$，它满足公理 1～6，而且具有性质：

1. 若存在 $J_1>\bar{J}_1,J_2>\bar{J}_2$，则

$$(J_1^*-\bar{J}_1)(J_2^*-\bar{J}_2)=\max(J_1-\bar{J}_1)(J_2-\bar{J}_2).$$

2. 若不存在 $J_1>\bar{J}_1$，则 $J_2^*=\max J_2$；若不存在 $J_2>\bar{J}_2$，则 $J_1^*=\max J_1$.

纳什的结果尽管完美，但他的公理并非无懈可击(特别是公理 4). 在实践中已形成了许多在某种

意义上更富理性或更富技巧的协商和谈判的策略.

谈判理论(negotiation theory) 即"协商理论".

两人零和对策(two person zero-sum game) 一种最基本的完全对抗情形的对策模型.局中人数为二,且任何局势下两人的赢得总数恒为零(或任何常数)的对策问题.这时"你之所得即为我之所失",所以双方的利益是完全对抗的.这种对策在政治、军事、企业竞争等敌对行为中有重要的应用,得到普遍重视和深入的研究.根据两人可以选择的策略数为有限或无限,对策问题亦分别称为两人有限或无限对策.其中,前者可以用一个支付矩阵来描述,故又称矩阵对策.矩阵的元素 J_{ij} 表示当甲方选策略 i 而乙方选策略 j 时,甲方将得到 J_{ij} 的支付,而乙方将损失 J_{ij}.两人零和对策最重要的概念是鞍点.在矩阵对策中,若策略对 (i^*,j^*) 对任何 i,j 满足

$$J_{ij^*}\leqslant J_{i^*j^*}\leqslant J_{i^*j},$$

则显然双方都不愿意偏离此点.因此只要鞍点存在,它就可作为两人零和对策的一个平衡解.但即使在最简单的情形鞍点也可能不存在,对策论的最重要成果之一就是引入混策略的概念,并证明了混策略的鞍点解一定存在.所谓混策略就是对策人可选方案的一种概率分布,在对策中他按此分布随机地抽取并使用某一策略.这时,双方的得或失亦相应地用双方混策略的联合分布所决定的支付函数的数学期望来决定.与此相对,原来非随机方式选取的策略就称为纯策略.关于混策略的研究是很有实际意义的,在军事上的搜索问题或动物追捕过程中,其规避行为经常是随机的曲折运动,而不可能是个预先设计好的行为.两人零和对策可以推广到微分对策的情形,并具有相似的结论.

鞍点解(saddle point solution) 见"两人零和对策".

纯策略(pure strategy) 见"两人零和对策".

混策略(mixed strategy) 见"两人零和对策".

非零和对策(non-zero-sum game) 一类非完全对抗的对策问题.局中人之赢得总和不恒为零(或任何一个常值)的对策问题.由于这时既可能通过协商、合作增加总的收益,导致各人均有所得,从而形成多人合作对策(参见"多人合作对策");也可能由于互不信任,无法合作,各取自己的个体优超策略,而导致总体收益下降,即所谓纳什平衡解可能导致囚犯悖论或社会悖论(参见"纳什平衡"、"囚犯悖论"、"社会悖论")的现象.在合理分配方案的研究中,纳什(Nash,J.F.)还发展了一套协商和仲裁的理论.这一大类对策问题在市场经济、国际合作等多方面有重要的应用,是当前对策论研究最活跃的领域,并多次获诺贝尔经济学奖.

纳什平衡(Nash equilibrium) 非零和对策的一种非合作平衡解.其特征为任一局中人单独偏离该平衡解均不会给他带来好处.具体形式化描述如下:设 N 个局中人的决策分别为 $u_1,u_2,\cdots,u_N$,其所得分别为 $J_1(u_1,u_2,\cdots,u_N),\cdots,J_N(u_1,u_2,\cdots,u_N)$.称决策组 $(u_1^*,u_2^*,\cdots,u_N^*)$ 为纳什平衡是指对任何 i,均有

$$J_i(u_1^*,u_2^*,\cdots,u_N^*)\geqslant J_i(u_1^*,u_2^*,\cdots,u_{i-1}^*,u_i,u_{i+1}^*,\cdots,u_N^*).$$

此不等式右端只有 u_i 改为可取任何值.该不等式表明,u_i 单独偏离 u_i^* 不会使 J_i 增大,从而该平衡解具有某种稳定性.但这并不排除若干局中人协同行动同时偏离该平衡点可以给各人均带来好处.一个著名的简单例子是所谓囚犯悖论(参见"囚犯悖论"),它可由其支付矩阵描述.其中各支付函数 (a,b) 表示 DM_1 所得为 a,DM_2 所得为 b.显然决策对 (D,D) 对应的支付 $(-5,-5)$ 为纳什平衡,因为任何人单方将决策改变为 C 将使他自己得到 -10 的结果.尽管两人同时改变决策可使他们均得到较好的结果 -1,只是由于他们的非合作性,这个较好结果是不稳定的,通常难以实现.同样的模型还可用来描述军备竞赛、对紧缺公有资源的掠夺、公益事业投资等问题.这种只顾及个人局部利益而导致对全局不利结果的现象,在社会生活中是较普遍的,这种现象称为社会悖论,因此纳什平衡解的研究对改进社会群体的管理有重要的现实意义(参见"社会悖论").

囚犯悖论(prisoner dilemma) 阐明纳什平衡概念的一个著名例子.设两名囚犯嫌涉参与某罪行而受审.他们商定合作顽抗,但被告知以下政策:若两人均招认(不合作,决策 D),则罪名成立,各判刑五年;若有一个招认(决策 D)而另一人顽抗(决策 C),则招认者立功释放,而顽抗者罪名成立,重判十年.但若两人均顽抗(决策 C),则可能由于证据不足而只能各轻判一年.以上各可能性可以用支付矩阵描述如下表.

囚犯悖论的支付矩阵

支付函数		囚犯 1 之决策	
		D	C
囚犯 2 之决策	D	$(-5,-5)$	$(-10,0)$
	C	$(0,-10)$	$(-1,-1)$

这个例子的纳什平衡解是 $(-5,-5)$,而合作最优解为 $(-1,-1)$,直观地说明了各人均用其优超决策 D 将导致较坏的结果,从而指出了纳什平衡的意义和缺陷.

社会悖论(social dilemma)　关于人类社会自相矛盾行为的一种悖论.特别是纳什平衡与囚犯悖论在多决策人的社会环境中的应用研究,其特点为各决策人均采用个人优超策略却导致对大家均不利的结果.在人口控制、军备竞赛、资源环境保护、市场竞争等许多重大现实问题都会达到这类统称为社会悖论的现象,从而它已成为现代决策和博弈研究的一个活跃领域.

研究表明,社会悖论的类型除了囚犯悖论,还有衰种悖论和信任悖论,它们可分别用来描述囤积和抢购及体育中的兴奋剂问题等现象.此外,还有反映眼前和长远利益之间矛盾的时间悖论,可用来描述吸毒成瘾等现象.阿克赛罗德(Axelrod,R.)还设计了一种多次重复囚犯悖论的模型,并利用计算机随机模拟来评估人们在和各种具不同"处世策略"的人随机相遇的社会环境中的累计效益,他向世界上著名博弈论专家征集了几十种"处世策略"进行试验.有趣的是由著名系统论和博弈论专家拉波波特(Rapoport,A.)提出的"一报还一报"策略一直高居榜首.这种策略的特点可简单概括为"诚恳待人,但遇到欺骗时应给以有限度的报复",其寓意也是深长的.

社会悖论的研究和教育对于制订社会公约、规范公民行为有重要的启发和参考价值,也是社会系统管理和信息经济研究的一个重要方向.

衰种悖论(chicken dilemma)　见"社会悖论".

信任悖论(trust dilemma)　见"社会悖论".

时间悖论(time dilemma)　见"社会悖论".

军备竞赛模型(arm race model)　一种博弈模型.用博弈论研究军备竞赛的一种模型.包括反映双方非合作导致囚犯悖论型纳什均衡的静态模型,以及描述双方对抗、消耗,渐至疲劳而达到某种均衡的动态过程.前者假设双方均有扩军和裁军两种选择,若双方扩军导致费用增加,双方均不如同时裁军更为有利.但若某方单独扩军,它将得到更大的政治和军事的好处.因此双方扩军是纳什均衡,导致双方均不大有利的结果.该问题的研究将有助于认识这一悖论,形成对全局更为有利的结局.后者则包括曾由理查森(Richardson,L. F.)提出的一种微分方程模型

$$\dot{x} = -\alpha x + ky + g,$$
$$\dot{y} = lx - \beta y + h.$$

式中 x 和 y 分别为双方军备水平;k 与 l 称"防御"系数,表示对对方扩军的反应强度;而 α 和 β 为"疲劳"系数,反映对扩军费用增长的厌倦程度.这是一种状态变量均只取非负值的正系统.已经证明,当 $\alpha\beta > kl$,即厌倦程度更大时,系统有双方均为非负值的平衡点,而且该平衡点是稳定的.这与常识亦相符.这一模型还可推广到多国军备竞赛的一般情形.

主从对策(leader-follower game)　亦称斯塔克伯格对策.一种多级递阶决策问题,由经济学家斯塔克伯格(Von Stackelberg)提出.设 L 为上级(主方),其决策变量为 x_L,信息为 z_L,策略为 $x_L = \gamma_L(z_L)$,而目标函数为 $K_L(x_L,x_F)$.而下级 F(从方)的相应变量亦类似地可表示为 x_F, z_F, $x_F = \gamma_F(z_F)$ 和 $K_F(x_L,x_F)$.主从对策问题就是主方应如何选择并宣布他的策略 γ_L,使下级在此前提下为优化自己的 K_F 而执行的 γ_F 最符合主方的利益.在较简单情形下,若上级直接宣布 x_L 的值,而下级可根据 $z_F = x_L$ 来选择,

$$x_F = \gamma_F(x_L) = \arg\max_{x_F} K_F(x_L,x_F),$$

这就是从方的理性响应.由于主方可以预见到这一结果,问题将归结为选择 x_L,使 $K_L = K_L(x_L, \gamma_F(x_L))$ 达到最大.由此可见,主从对策实际上是对从方行为的间接控制问题.

这种提法可以推广到多阶段的动态的情形.设主从双方共同控制一个动态系统 $\dot{x} = f(x, u_L, u_F, t)$, $t \in [0,T]$,这时 u_L 和 u_F 均为时间函数,相应的目标函数则为

$$K_i = \varphi_i[x(T)] + \int_0^T \psi_i(x,u_L,u_F,t)\mathrm{d}t$$
$$(i = L,F).$$

这时问题仍然是主方如何选择最优的 u_L^*,使从方以此代入自己的问题后求解 K_F 为最大所得的 u_F^* 值恰好能使主方得到最好的 K_L.由于这里涉及一种带有未知函数 u_L 的最优控制问题族对 u_F 的求解问题,理论上有很大难度,其结果甚至还与主方拥有怎样的动态信息有关,即 u_L 是开环式的时间函数 $u_L(t)$,状态反馈式的 $u_L(x)$,还是带记忆的反馈

$$u_L(t) = \gamma_L[x(\tau), \tau \leqslant t],$$

这都会带来不同的后果.当主方直接根据 $z_L = u_F$ 来进行决策,亦即制订"政策" $u_L = \gamma_L(u_F)$ 来控制从方的行为时,即是所谓激励控制问题(参见"激励控制").

主从控制的思想可用于多级管理结构中为上级制定恰当的策略来对下级的行为进行间接的、宏观的调控,有重要的实际意义.

斯塔克伯格对策(Stackelberg game)　即"主从对策".

激励控制(incentive control)　多级递阶系统的一种决策和控制方法.在多级递阶决策问题中,上级(主方)决策人可以用主从对策的方法来控制下级(从方)的行为.上级只要宣布适当的"政策" $u_L = \gamma_L(u_F)$,通过下级自身的优化行为进行诱导和激励,就可以达到控制整个系统行为的目的,因此称为激

励控制. 激励控制的理论包括可控区的界定、缺乏下级私有信息时的估计策略，以及各种类型问题的激励控制设计方法. 已有的应用包括企业定额管理、资源分配等方面.

具非对称信息对策问题(game problem with asymmetry information) 各局中人掌握信息不同时的对策问题. 它在现代经济活动中有重要应用. 例如，最著名的"主管(亦称委托)和代理人"对策中，主管掌握更多权力而代理人具较多私有信息. 后者可能通过隐瞒、虚报等来为自己谋利. 在企业管理和市场竞争中有大量例子. 英国经济学家梅里斯(Mirrless, J.)等关于这一问题的研究获得了 1996 年诺贝尔经济学奖. 这些工作已成为信息经济学的重要理论基础.

微分对策(differential game) 亦称动态对策. 一种考虑动态决策过程的动态的对策理论. 指各决策(局中)人均将各自的控制作用施加于同一(连续或离散时间的)动态系统并由此确定各自的品质指标的情形. 其中各人还可对系统进行观测，获取实时动态信息来修改自己的策略，对别人的行为进行鼓励或报复等灵活反应. 微分对策问题提法十分多样，问题也极为复杂. 微分对策最早由艾萨克斯(Issacs, R.)等人于 1965 年提出，主要用于研究以空战、格斗等为背景的两人零和问题. 设动态系统 $\dot{x}=f(x,u,v,t)$ 中 u 和 v 分别为两决策人的控制函数，他们的得失由品质指标

$$J(u,v)=\Phi[x(T),T]+\int_{t_0}^{T}L(x,u,v,t)\mathrm{d}t$$

来描述，u 方希望 $J\to\min$，而 v 方希望 $J\to\max$. 对于 u,v 均为时间函数的情形，可以有推广的"双方极大值原理"，它指出鞍点策略 $u^*(t)$ 和 $v^*(t)$，$J(u^*,v)\leqslant J(u^*,v^*)\leqslant J(u,v^*)$，应满足的必要条件称为哈密顿-雅可比-贝尔曼-艾萨克斯条件，即对任何 $t\in[t_0,T]$，均有

$$\begin{aligned}&\max_v\min_u H(\lambda,x,u,v,t)\\&=\min_u\max_v H(\lambda,x,u,v,t)\\&=H(\lambda,x,u^*,v^*,t).\end{aligned}$$

其中哈密顿函数

$$H(\lambda,x,u,v,t)\triangleq\sum_{i=1}^{n}\lambda_i f_i(x,u,v,t),$$

而 $\lambda_i(t)$ 为协态变量，它满足

$$\dot{\lambda}_i=-\frac{\partial H}{\partial x_i}$$

及边界条件

$$\lambda_i(T)=-\left(\frac{\partial\Phi}{\partial x_i}+\mu\frac{\partial\Psi}{\partial x_i}\right)_T\quad(i=1,\cdots,n).$$

这类问题的可解性显然取决于函数 H 的鞍点存在与否，否则还要用到混策略的概念. 以上求解条件也可以用动态规划的思路来推导，并得到类似于贝尔曼方程的求解条件.

在捕获-逃逸的动态对策问题中，还经常会遇到诸如哪一方能最终达到目标之类的定性微分对策问题. 对此亦有许多理论研究. 例如，可以在状态空间中划分出可捕获区和可逃逸区，并说明它们与系统模型参数(诸如飞机的速度和转弯半径等)之间的关系，这些均有重要的实际参考价值. 对于动态非零和对策问题，包括动态纳什对策、动态主从对策，也都有许多研究. 虽然也可列出极大值原理型的条件，但往往导致非古典的、未知函数相互耦合的两组两点边值问题的求解，目前均难以处理. 各控制函数依信息结构不同，可以是开环式的时间函数、状态反馈形式或是带记忆的状态反馈(称为闭环解)，它们都会导致本质不同的结果. 对动态纳什对策，可以证明状态反馈解就是闭环解，它们都满足最优性原理，并可用动态规划求解. 但对动态主从对策这一结论不成立(参见"主从对策"). 由于这里涉及主方对从方的"错误"行动，应有足够的信息，及时发现，仍有机会给以惩罚，这里需要一些比较特殊的信息结构，以及 u_L 对 J_F 足够的影响力.

哈密顿-雅可比-贝尔曼-艾萨克斯条件(Hamilton-Jacobi-Bellman-Issacs condition) 简称 HJBI 条件. 零和动态(微分)对策求解的一个基本必要条件. 即对任何 $t\in[t_0,T]$，均有(参见"微分对策")

$$\begin{aligned}&\max_v\min_u H(\lambda,x,u,v,t)\\&=\min_u\max_v H(\lambda,x,u,v,t)\\&=H(\lambda,x,u^*,v^*,t).\end{aligned}$$

该方程解的存在性和计算方法即为零和动态对策的基本问题. 若此条件不满足，则应研究其混策略解是否存在及求解问题.

动态对策问题的开环解(open-loop solution for dynamic game problem) 动态对策的一种特殊解. 指动态对策具有开环信息结构时的解. 这时，对策双方均无任何实时信息，只能依据预知的各种知识. 因此，双方必须预先给出整个对策期间各自的决策值 $u(t)$ 和 $v(t)$，$t\in[0,T]$，不可中途进行任何修改. 这种问题可以看成在函数空间 $U=\{u(t),t\in[0,t]\}$ 和 $V=\{v(t),t\in[0,T]\}$ 中的静态对策问题.

动态对策问题的反馈解(feed-back solution for dynamic game problem) 动态对策的一种特殊解. 指动态对策具有瞬时状态信息反馈情形的解. 这时，双方可取策略形式为

$$u(t)=F[x(t)],\quad v(t)=G[x(t)].$$

已有例子表明，该信息结构下的最优解不能用动态规划技术来求得，而且和信息更完全情形的"闭环

解"有不同的结果(参见"动态对策问题的闭环解").

动态对策问题的闭环解(closed-loop solution for dynamic game problem) 动态对策问题的一种特殊解.指动态对策具有记忆状态信息反馈情形的解.这时双方可取策略形式为

$$u(x)=F[x(\tau),\tau\leqslant t],\ v(t)=G[x(\tau),\tau\leqslant t].$$

按具体形式不同,其中主方还可能比对手(从方)拥有更完全、更超前的信息,如主从对策(参见"主从对策")中决策 $u(t)$ 的主方可以由此导出从方的决策变量 $u(\tau),\tau\geqslant t$,并及时做出反应.这种信息优先使主方可以用激励策略奖励或惩罚从方,使之采取更为合作的策略(参见"激励控制").对这种更为复杂的控制策略已有若干可行的求解方法.

多人合作对策(multi-person cooperative game) 研究多人决策合作、结盟和利益分配的对策问题.多个局中人可能具有不同决策变量和信息,并追求各自不同目标.通常易于描述和处理的是具完全信息的一次性行动的情形.这时,又可按局中人的合作性分为非合作对策和合作对策.非合作情形导致纳什均衡解等概念,并如囚犯(社会)悖论现象所指出,常常带来各方都不利的结果.另一类则考虑局中人合作可能带来的收益,研究如何激励社会各方的合作行为.研究这种情形的出发点是结盟和特征函数的概念,它指明了一旦局中人集 N 的子集 S 中各人结盟合作时,他们至少可保证得到的总收益,记为 $V(S)$;然后为保证这种合作是稳定的,他们还必须能找到一个合理的分配方案.首先,按 $V(S)$ 的定义,可以约定空集的特征函数为零,$V(\phi)=0$.其次,$V(S)$应为下列两人零和问题的解:以 S 的成员结盟为一方力图使其总收益最大,而以 S 以外成员结盟为另一方并力图使 S 的总收益为最小.此外,若 S 和 T 为任意两个不相交的对局人子集,则形成结盟的首要条件自然是

$$V(S\cup T)\geqslant V(S)+V(T).$$

具这种性质的特征函数称为超可加的.

对 n 人问题,各种结盟的可能性有 2^n 之多,因此用特征函数的描述仍过于复杂.有若干较简单的例子可以用来说明社会生活中可以遇到的结盟和利益分配问题.为了进一步研究结盟的可行性,须研究各种分配方案的合理性,并从中决定其真正实现的可能.

芯(core) 多人合作对策中结盟 S 的特征函数.$V(S)$的一类分配方案,它不被其他分配方案所优超,通常记之为 $C(V)$,它是多人合作对策的一种解概念.显然,芯中的分配方案是合理、可行的,但芯的要求较高,它经常是空集.

稳定集(stable set) 多人合作对策的一种解集.满足下列条件的分配方案构成的集合$S(V)$称为稳定集:

1. 若分配方案 $x,y\in S(V)$,则 x 不能优超 y,y 也不能优超 x.

2. 若 $y\notin S(V)$,则存在 $x\in S(V)$,使得 $x>y$.

这里,条件 1 可防止不同部分对策人利益冲突引起分裂,称为内稳定条件;而条件 2 则强调 $S(V)$ 以外的分配方案可不必考虑,称为外稳定条件.这些条件均比芯的要求弱,因而有 $C(V)\subseteq S(V)$,稳定集条件较弱,但仍不能保证它必存在.为此人们又提出了核的概念,它永远存在惟一而且只要芯存在,它必为芯的一部分.为计算核,先要计算各子集 S_i 对于分配向量 x 的余量 $e(S_i,x)=V(S_i)-x_i$,亦即 S_i 结盟时所得比按 x 分配时增加的好处,而好的分配向量应使这种余量尽可能小.通过比较复杂的比较程序可得到余量最小的分配,向量就称为多人对策特征函数 $V(S)$和给定的可取分配向量集 $Z=\{x\}$的核.其他的重要概念,如夏普利值$\Phi_i(V)$,反映决策人 i 在博弈 V 中可预期获得的收益.在诸如联合国安理会投票的模型中,它反映了各成员(常务和非常务理事国)间实质上的权力分配,有重要实际意义.

此外还有核心、支付模式等概念,均探讨结盟的可能性及其分配方案.所有这些在社会经济管理中,如社会权利和义务的分配,均有重要应用.

核心(kernel) 见"稳定集".

夏普利值(Shapley value) 见"稳定集".

支付模式(payoff pattern) 见"稳定集".

圣彼得堡悖论(St. Petersburg dilemma) 有关效用理论和决策行为的一个悖论.它表明:用数学期望来评价具有不确定收益的决策的品质优劣,可能导致不合理的结果,由尼古拉第二·伯努利(Bernoulli, Nicolaus Ⅱ)于 1728 年提出.假定某人重复抛一对称硬币直至出现正面为止,若抛币次数为 n,则得 2^n 元奖励.由于该事件出现概率为 $1/2^n$,因此其期望收益为

$$\sum_{n=1}^{\infty}2^n(1/2^n)=\infty,$$

即大于任意常数 C.但事实上人们不愿付出较大的代价 C 来参加这一赌博.这一矛盾使人们认识到,必须用与期望收益非线性相关的效用函数来作为决策依据,由此导致美籍匈牙利数学家冯·诺伊曼(von Neumann, J.)与莫根施特恩(Morgensten, O.)的主观期望效用理论的建立.

艾利斯伯格悖论(Ellisberg dilemma) 有关效用理论和决策行为的一个悖论.它表明主观期望效用理论亦存在不合理性,由艾利斯伯格(Ellisberg)于 1961 年提出.设两桶 A 和 B 中各装红球和黑球共 100 粒,A 桶中红黑各半,B 桶中红黑比例未知,今有四种行动导致各自后果:AR——由 A 桶取球,

猜为红，若对则得 50 元，否则得 0；AB——同 AR，但猜球为黑；BR，BB 则为以 B 桶取球，其他规则均相同. 调查大多数应试者的偏好顺序为

$$AR \propto AB > BR \propto BB.$$

但可以证明，这与主观期望效用理论相违背.

控制论系统

控制论系统(cybernetic system)　基于更广泛的控制论原理建造并具有更强适应性的控制系统. 由于基于反馈原理的自动调节和控制的理论方法已比较成熟，人们进一步关注如何实现具有更复杂的适应、学习、修复、重组、自组织和繁殖等更接近于生物行为的复杂系统，即更高层次上的自动机器. 其具体功能可包括以下若干方面：

1. 通过量测周围环境和系统内部状况的变化，调整控制机构以保持系统的稳定性和控制性能.

2. 自动检测系统故障并进行修复和重组.

3. 积累运行数据，自动归纳、学习，以保持预定的功能.

4. 通过搜索或对人为施加刺激的反应，自动搜寻并保持系统的最佳运行状况.

5. 具有一定的自我演化能力，不断改进系统功能学习、存贮或自我复制优良的控制机制.

6. 良好的人机交互界面，具有博弈游戏或协助人类从事各种活动的能力等，由于计算机通信和控制技术的迅速发展，具有以上复杂功能的机器已逐步得以实现.

7. 具有多种感觉器官和拟人的执行机构，从而可以替代人们从事某些较复杂的脑力和体力活动的智能机器人系统.

这类系统在理论上可形式化地将其定义为一个六元组

$$S \triangleq \langle I, \Omega, \alpha, Z, \gamma, Q \rangle,$$

其中：I 为输入集或输入函数，它来自环境 E，可以是随机的或特定的时间函数；Ω 为系统作用于环境 E 的输出集或输出函数，亦可为变化的或特定的时间函数；α 为前向传递映射，其值域确定了系统行为的定义域；Z 为品质评估映射或函数，它对 I 和 Ω 进行运算并指出系统对其目标或约束的行为特征；γ 是系统的参数映射，它把系统品质测度映射为适当的系统参数，从而通常可指出系统改进其未来行为的方式；Q 为状态集或状态函数，其值指明了系统行为域的某一子集. 这些定义均比较宽松，具体内容亦多种多样. 其主要目的是涵盖人类研究更为高级、复杂的自动化系统的若干发展方向.

自寻最佳系统(self-optimizing system)　亦称极值控制系统. 可自动搜索和趋近于最佳运行状态的系统. 它能自动试探和调整参数，使系统趋近并保持于最佳工作点. 其工作原理为：每次使被调整参数 θ 增加或减小一个小量 $\Delta\theta$，量测并评估其导致系统品质的增量

$$\Delta J = J(\theta + \Delta\theta) - J(\theta),$$

依其大小和符号决定是否及向何方调整 θ 值，最终力求使 θ 保持在最佳值附近(为了搜索，它不可能总保持于最佳值). 这种方法特别适用于机理不明或过于复杂的系统. 最早由美国德拉帕(Draper, C. S.)和华裔控制论专家李耀滋于 1940 年提出，并用于发动机运行参数的优化问题. 其后苏联、中国都有人进行研究和应用. 其主要研究问题是极值参数的搜索和保持策略，使在随机干扰条件下以较小的搜索代价达到系统优化的目的.

极值控制系统(extreme control system)　即“自寻最佳系统”.

自动控制系统理论(theory of automatic control system)　研究自动控制系统分析、设计和实际应用的理论. 自动控制系统是一类力求以尽可能少的人类干预实现尽可能多的自动监视、检测、调节和控制作用以达到预期技术要求的人造系统. 自动控制理论则研究这类系统的构思、设计、性能、分析，乃至实施和运行的原理和技术.

尽管有许多史前的例子，瓦特蒸汽机上的飞轮调节器应属最早自觉地完善设计及成功应用的第一个自动调节系统，并在工业革命中起到重要作用. 麦克斯韦(Maxwell, J. C.)于 1868 年发表其稳定性分析的论文，应属最早的自动控制理论成果.

自动控制系统的标称行为须用指令来设置，它可以是逻辑开关变量或者连续变化的时间过程. 前者靠逻辑验证来保证其正确性，后者则利用检测获取的输出信号，经过加工处理反馈到系统输入端进行调整，以减小系统的误差.

几乎所有的误差均来自不确定性，包括系统内部特性的误差、飘移，外界环境的变化和干扰. 这些误差大多数都可以用反馈技术加以削弱或消除. 稳定性可以保证消除初始值误差，而其中频率域分析方法又由其对系统特性(阶数、参数)不敏感而得到更广泛的工程应用. 随机系统控制理论研究随机信号和噪声的影响和消除，以克服统计的不确定性. 最优控制理论研究某些宏观品质指标的优化，而略去许多不确定的细节. 鲁棒控制旨在设计一种在整个系统特性变化范围均能正常工作的系统. 自适应、自校正控制则可按系统特性变化随时跟踪调节控制参数，以保证其正常工作. 模糊控制处理具“模糊性”的不确定因素. 自寻最佳和学习控制均采用试探并分析已有反应特性来改进系统品质. 利用各种人工智能或软计算方法(包括神经元网络、启发式规则、演

化算法等)的智能控制,可模仿人类处理复杂不确定问题的经验和技巧.目前,非线性控制实际上仅仅触及千变万化的非线性现象的一些特殊情形.从而可以认为,整个控制理论基本上都是围绕克服不确定性这一目标而展开的.

计算机和逻辑控制技术的迅速发展和广泛应用引出了离散事件和混杂系统的控制理论,尽管它的理论基础可以回溯到1938年创立的继电开关电路的分析和逻辑设计理论,它在20世纪80年代的再次繁荣仍应归功于技术发展的需求和推动.离散事件和混杂系统处理的不仅仅是开关型和连续变量两种控制范式的结合和应用,也是复杂系统宏微观层次定性与定量运动形态及其控制手段的综合集成.这里还要强调,它们实际上也符合智能控制的基本思路,如逻辑推理、知识和规则、启发式搜索等,而且这些原则最后也都将归结为描述定性行为的离散事件过程,并可用严格的理论和先进的技术手段来实现.

综上所述,自动控制系统的发展趋势可以概括为多方面的复杂性的综合:

1.控制对象复杂.包括大而复杂的多层次结构,多种运动形态、关系复杂、规律不明等.

2.环境条件复杂.未知、多变的各种干扰须用不同方法描述和处理.

3.控制手段复杂.包括大型计算和通信设备,多种传感技术,人和机器的交互等.

4.控制问题目标要求,可用资源的代价和约束条件也是复杂的,它常常是更大系统范围综合平衡的结果.

大系统分散控制(large-scale system decentralized control)　将控制功能分散化的一种控制途径.每个控制站只采集或接收部分信息,包括其所在子系统的反馈信息及某些其他子系统传送的信息,同时它的控制作用只作用于该子系统本身,但可通过互联通道间接地影响其他子系统.控制目的仍然是使系统稳定,按某种确定性或随机的品质指标最优,保证系统在出现模型误差或有较大干扰时具有鲁棒性质等.这里仅以线性系统情形为例,介绍其具体形式.设系统方程为

$$\dot{x} = Ax + \sum_{i=1}^{N} B_i u_i,$$
$$y_i = C_i x,$$

亦即系统矩阵 B 和 C 均具有分块对角形式,其中第 i 个控制站的控制作用为 u_i,对应观测值为 y_i $(i=1,2,\cdots,N)$.与常规的集中控制模式相对应,分散控制也有几种基本的反馈方式,诸如:

1.分散状态反馈:

$$u_i = K_i x_i, \quad \sum_{i=1}^{N} n_i = n,$$

式中,x_i 为第 i 个子系统的状态,其维数为 n_i.

2.分散输出反馈:

$$u_i = K_i y_i, \quad \sum_{i=1}^{N} r_i = r,$$

式中,y_i 为第 i 个子系统的输出,其维数为 r_i.

3.分散动态反馈:

$$\dot{z}_i = F_i z_i + G_i y_i, \quad u_i = H_i z_i + K_i y_i,$$

式中,z_i 为动态补偿器的状态,其维数一般不超过 n_i.

采用分散控制的合理性来自系统结构本身的分散特性、子系统的相对孤立性,以及其间传送信息带来的噪声、时延和代价等.因此,虽然理论上它肯定达不到完全信息反馈控制的品质,但权衡代价与得失,在现有技术条件下分散控制仍是大系统控制的现实可行的途径.待研究的问题包括:判断是否可用分散控制来镇定整个系统,分散控制导致的次优品质的评估,以及具随机干扰和分散信息情形的组队优化控制等.

大系统分散随机控制(large-scale system decentralized stochastic control)　具有随机干扰的大系统控制问题.它包括基于分散化量测、估计而进行的分散控制和随机优化.以离散时间情形为例,其数学模型为

$$x(k+1) = Ax(k) + \sum_{i=1}^{N} B_i u_i(k) + \xi(k),$$

其中 x 为 n 维状态向量,$u_i(k)$是第 i 控制站 k 时刻的 m_i 维控制;$\xi(k)$为零均值、方差为 $V(k)$的白噪声序列;控制站 i 的量测结果为 $y_i(k)=C_i x(k)+\eta_i(k)$,噪声 $\eta_i(k)$亦设为零均值、方差为 $W_i(k)$的白噪声序列;其最优控制问题为:求分散控制作用 $(u_i(k)(i=1,2,\cdots,N,k=1,2,\cdots,K)$,使下列费用函数极小化:

$$J = E\left\{\frac{1}{2}\sum_{k=1}^{K}\left[x^{\mathrm{T}}(k)Q(k)x(k) + \sum_{i=1}^{N} u_i^{\mathrm{T}}(k)R_i u_i(k)\right]\right\} \to \min,$$

其中 $Q(k)$,R_i 均为适当维数的正定矩阵.

这一问题的求解与系统的信息结构有密切的关系.常见的信息结构包括:具有完全信息的"非分散解",控制站之间具通信延迟的"一步延迟共享信息结构"等.威腾豪森(Witsenhausen,S.)曾举出一个非古典信息结构的反例表明其不可解性,它又与多人组队决策问题有密切的关系(参见"组队决策").

大系统关联稳定性(large-scale system connective stability)　由关联结构确定的大系统稳定性.

即当大系统的某些子系统间的关联项断开时，原有的稳定性质是否仍能保持的问题. 子系统间这种关联关系可用关联矩阵 $E=(e_{ij})$ 来描述，$e_{ij}=1$ 或 0，分别表示子系统 j 对子系统 i 有无影响. 若大系统 S 的平衡点 $x=0$ 在所有可能的关联阵 E 时均为稳定，则称系统 S 的零解为关联稳定. 进一步的要求还可类似地导致一致关联稳定、全局一致关联稳定、全局指数关联稳定等概念.

大系统固定模(fixed mode of large-scale system) 大系统分散控制结构所决定的无法经由分散控制加以改变的运动模态，它是传统集中控制系统理论中不可控、不可观模态概念的推广. 在大系统分散控制的基本模式

$$\dot{x}=Ax+\sum_{i=1}^{N}B_iu_i,$$

$$y_i=c_ix,\quad u_i=K_ix_i$$

中(参见"大系统分散控制")，它是由任选 K_i 所能实现的本征值的交集，从而它与 K_i 无关，是不可经控制设计而加以改变的. 大系统可用分散控制加以镇定的充分必要条件之一就是系统的固定模 $\Lambda(A,B,C)$ 均稳定. B 为分块矩阵$(B_1,B_2,\cdots,B_N)$，C 为分块矩阵 $C=[C_1^{\mathrm{T}},C_2^{\mathrm{T}},\cdots,C_N^{\mathrm{T}}]^{\mathrm{T}}$ (T 表示矩阵转置). 单纯从矩阵 B_i 或 C_j 的零元和非零元的分布情况有时也可以判定相应矩阵的秩，从而可知固定模的存在和振型. 这种固定模称为结构固定模.

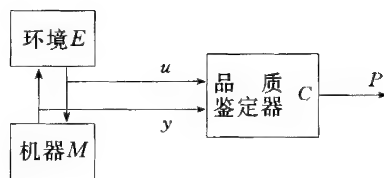
结构固定模(structure fixed module) 见"大系统固定模".

反馈系统(feedback system) 具有闭环信息通道的系统. 将系统的后果或输出信息采集、处理，然后送回输入端并据此调整系统行为的系统. 由于信息流通构成闭合环路，它亦称为闭环系统. 反馈作用常用于检测信号偏差及对象特性的变化，并依此来控制系统行为及消除误差. 它又被称为反馈控制或按误差控制的系统. 理论上可证明这类系统具有较好的稳定性、抗干扰性和对系统模型变化的不敏感性，从而成为自动控制系统的基本结构.

与反馈控制相对应的是开环(前馈)控制系统. 由于它具有响应迅速等优点，亦被用于控制系统中，与反馈共同使用构成组合控制系统.

闭环系统(closed loop system) 见"反馈系统".

自适应系统(adaptive system) 能调整自身以适应环境变化的系统. 这种对环境变化不敏感的系统，可以用机器、环境、品质鉴定器的关系图来说明. 如图所示，环境与机器相互作用，并由品质鉴定器给出品质函数值 P. 若 W 表示容许品质函数集，对任何 $u\in U$，均有 $P(u)\in W$，则称该系统对 U 和 W 为



自适应的.

在技术上，可以用各种自适应控制的方法构造所需的自适应系统. 更为一般的自适应定义来自生物系统的概念. 例如，若系统能在变化的环境中将其本质变量保持在一定的生理学限度内，则其行为方式就是自适应的. 又如，若系统某些部分受到损害时，这种损害的效应能逐步隐蔽起来，从而系统的性能可保持在容许范围内，则此系统就是自适应的.

工作于随机干扰下的系统，其控制作用将同时起到试探激励的作用，以加快对环境参数和系统特性的统计估值过程，有利于改进控制品质. 这种由费尔得鲍姆(Feldbaum, A. A.)提出的二重控制的概念，也被称为随机自适应控制. 自适应控制系统必须在运行过程中不断对控制对象和环境的特性进行测量和辨识，并据此修改控制器的参数或结构. 这里辨识途径可以是显式的辨识算法，也可以是利用参考模型之类的隐式方法. 前者主要是估计某些关键参数，故又称参数自适应控制. 而后者则基于对系统品质的间接判断，故亦称为品质自适应控制.

随机自适应控制(stochastic adaptive control) 见"自适应系统".

参数自适应控制(parameter adaptive control) 见"自适应系统".

品质自适应控制(performance adaptive control) 见"自适应系统".

二重控制(dual control) 见"自适应系统".

自行镇定系统(homeostat) 模拟生理参数稳定机制的一种机电模型. 最早由阿什贝(Ashby, W. R.)提出. 其基本思想是：当系统由相空间任一点出发，其运动轨线达到指定区域边界时，系统参数将按预定的规则转变为一组新的值并使系统继续在区域内运转，直到它达到区域中某稳定点时这一过程才停止下来. 阿什贝曾给出一个包括四组线圈和磁铁的简单模型. 对比较简单的系统模型可以求出它确能自行镇定的概率，结果表明这一概率将随系统的复杂程度(例如阶数 n)增高而迅速下降，其数量级大体为 $1/2^n$. 这意味着搜索的时间亦将随 n 按指数率增加，这就使这种方法实际上难以实现. 为克服这一困难，阿什贝采用多级递阶的分散结构，分组逐次镇定，从而使自行镇定得以实现. 这种系统亦被称为多稳定系统.

自学习系统(self-learning system) 亦称学习

系统.模仿生物学习功能的系统.它能在系统运行过程中通过评估已有行为的正确性或优良度,自动修改系统结构或参数以改进自身品质的系统.与自适应系统不同之处在于:经学习而得到的改进可以保存并固定在系统结构之中,从而较易于实现,并可作为自动设计或调整的一种办法.

学习方法可以分为两大类:一类是有导师监督对优良方案加以强化的学习,按预设的指标来评估品质并指导系统的改进;另一类则是无导师监督的学习,这时须要用试探、搜索等办法来探索改进的途径.由于人工神经网络、演化计算等高速并行处理技术的发展,无导师监督学习方法也已得到成功应用.

学习系统(learning system) 即"自学习系统".

自繁殖系统(self-reproduction system) 具有自我复制能力的系统.最早由美籍匈牙利数学家冯·诺伊曼(von Neumann,J.)于 1948 年提出.近代计算机病毒和多智能体的行为特征表明它们都具有自繁殖的功能.对生物遗传密码的破译将有助于揭示繁殖和遗传的奥秘,以建立功能更强、更符合人类需要的自繁殖系统.

自修复系统(self-repairing system) 具有自修复能力的系统.其功能包括故障检测和诊断,按设计的策略排除故障或进行系统重构,从而保持系统正常的基本功能.对于重要的大型核反应堆、航天飞行器等技术系统,这些都是最基本的要求.自修复系统涉及的主要技术包括故障诊断、系统冗余和重构、容错计算等.

经济控制论(economic cybernetics) 关于经济系统控制的理论,是控制论与经济学的交叉学科.它将控制论的原理和方法应用于经济活动及其管理过程,揭示经济系统调节和控制的一般规律和方法.当前主要工作是把 20 世纪 60 年代兴起的现代控制理论,主要是线性系统理论、最优控制理论等应用于经济活动分析,研究经济系统的稳定性、能控性、能观测性、鲁棒性、可靠性,以及反馈和优化控制等方法.由于对象的特殊性,经济控制论的基本概念,如输入、输出、状态、行为等都反映出经济系统的特殊内涵,系统的指令、决策及调节、控制职能中都包含人的因素,因而它又是系统科学与经济科学的交叉学科.

社会控制论(social-cybernetics) 将控制论原理用于社会问题的分支学科.它力求揭示社会系统管理和控制的一般规律和方法.但由于社会系统极度复杂,尚难以将现代控制理论的定量化方法直接用于研究社会活动.某些社会问题可以用人类决策或博弈论的方法来分析和解决.但总的说来,社会控制论仍然是一个有待开发的重要领域.

生物控制论(biological cybernetics) 控制论与生物学的边缘交叉学科,是控制论原理和方法在生物系统中的应用.它研究生命机体内调节控制和通信的一般规律和方法.现代生物控制论的主要奠基者维纳(Wiener,N.)在 20 世纪 40 年代发现并总结了生命机体和机器中的通信和控制的共同规律,建立了控制论这个边缘学科,并明确提出生物控制论的基本思想.但体系化了的生物控制论主要是 20 世纪 60 年代把控制理论的定量化方法引入生物控制问题研究而建立的.原则上说,生物系统的调节控制和信息处理存在于生命机体的各种层次上,但生物控制论主要研究的是器官、生理层次的调节控制,以及神经系统的信息处理.生物控制论研究已取得丰硕成果,有助于阐明生命机体的调控机理和感觉、记忆等认识规律,为医学工程和神经计算机研制提供设计思想.

复杂适应系统理论(complex adaptive system theory) 简称 CAS 理论.由具有学习、适应能力的个体所组成系统的整体演化机制和过程的理论.为美国圣菲学派的霍兰(Holland,J.)等人于 20 世纪 80 年代后期提出.该理论的中心思想是系统的适应行为导致世界复杂性的形成.其基本概念可概括为:

1. 主体,指具有自身目标、结构和行为方式并能根据环境变化调整自身目标、结构和行为从而主动适应环境的实体.大量这样的实体的相互作用,是系统整体得以生存和演化的动力来源.

2. 聚合,指上述主体在一定条件下相互集结、演化、粘合形成更大个体的过程和结果,聚合不是简单的合并或吞并,而是更高层次的新型个体的创生.

3. 非线性,指上述主体的行为和特性呈现非线性特征,并存在多重复杂反馈关系.这是聚合演化的必要条件.

4. 多样性,指主体通过相互作用和适应性过程,导致它们沿着不同方向、不同路线发展变化,从而形成个体类型的多样化,这是 CAS 复杂性的重要来源.

5. 流,指主体之间及它们与环境之间的物质流、能量流和信息流,最重要的是信息流.这些流的性态,如它们是否畅通、周转速度如何等,均将直接影响系统的演化.

6. 标识,指个体相互识别及系统选择个体所依据的个体特征,它们将用于实现信息的交流.

7. 内在模型,指主体学习和适应所依赖的内部机制,这种机制常具有层次、嵌套的结构.

8. 积木构件,指构成不同复杂适应系统的通用"建筑单元",不同系统的差别被认为主要是由这些"构件间"不同的相互作用和组合方式造成的.系统对环境的适应性演化就是通过改善和重组自己的

“积木构件”来实现的，一个层次上的构件通过适应、演化而聚集成更高层次的砖块.

目前，复杂适应系统研究的主要方法是计算机模拟，已开发了通用的SWARM软件平台，并已对社会经济、企业管理、生物和生态等系统问题进行了大量定性和定量的研究.

人机系统(man-machine system)　人与机器共存与合作的系统. 通常机器用来执行繁重的、可程序化的体力和脑力劳动，以减轻操作人员的负担. 根据人机交互层次的不同，可以分为三类：

1. 动力学层次. 主要研究机器如何与人类协调进行动力学控制之类问题，如驾驶舱仪表设置时如何考虑人的跟踪反应能力.

2. 技能层次. 即把人类操纵的技能和经验存入机器，提高系统的技术水平和品质，例如自动驾驶仪等.

3. 智能层次. 具有更高层的人机交互，人类只须确定系统目标、实现原则和约束条件等模糊因素，有时甚至还可计入疲劳、情感等因素，目前智能计算机研制中人机交互与和谐已成为最重要问题之一.

离散事件动态系统

离散事件动态系统(discrete event dynamic system)　简称DEDS. 描述离散、逻辑变量演化规律的动态系统. 其特点在于它的状态是离散逻辑变量，其演变则由某些外生、异步、突发的事件来触发. 通常其状态和事件各属于某些离散、有限的集合，从而上述演化机制可用类似于有限自动机的五元组来定义：

$$S = (Q, \Sigma, \delta, q_0, Q_m).$$

其中，状态 $q \in Q$ 的演化规则为由 $q = q_0$(初值)出发，若事件 $\sigma \in \Sigma$ 发生，则新状态按迁移律应为 $q' = \delta(q, \sigma)$，Q_m 为标识状态集，这些状态往往表示某种任务的完成或过程的终止. 以上是最简单、最基本的逻辑层的离散事件动态系统模型. 它的逻辑行为表现为由它给出的事件串的集合，称为由 S 生成的子语言 $L(S)$. Σ 中的事件有些是可控的，即当它出现时可以人为地阻止它发生，另一些则是不可控的. 通过阻止可控事件的发生，可以排除一些不希望出现的行为以满足人们的要求，这就是监控问题，最早由旺纳姆(Wonham, M.)和瑞梅基(Ramadge, P.)提出和研究.

离散事件动态系统的进一步研究涉及事件发生的种种机制. 较简单的是具确定性迁移时间的系统，即在每一状态，其下一个事件是按确定规则在确定时间发生. 对这类系统可以用极大代数或更一般的双子代数来描述(参见“极大代数”、“双子代数”). 在这种代数系中，系统的模型酷似传统控制理论中的线性系统方程，从而利用线性系统的已有框架可以建立起一套平行的理论. 当然，这时各对应的物理概念将有完全不同的解释.

对这类系统有两种控制途径. 一种是控制事件发生的时间而不改变逻辑顺序，这可用极大代数进行描述和求解. 另一种途径是控制事件发生的逻辑关系，这对应于实际系统中的路径规划、排序和调度问题. 由于系统的离散特性，这些问题常具有组合的计算复杂性，难以求解. 实用的常常是近似、次优或启发式的途径. 特别是有一种分布式反馈的实时调度策略，只依据局部的实时信息，用比较简单的规则就可以保证系统稳定运行，从而具较大实用价值. 由于用了反馈策略，还可使系统在不确定及变化的环境中工作.

更复杂一些的事件发生机制是随机性的. 这种系统可以用排队网络或马尔可夫链等来描述. 近来较多采用的还有广义半马尔可夫过程，它可以更全面地考虑各种因素，并易于计算机进行仿真.

以上各种模型在用于实际问题时都有计算复杂性或其他求解的困难. 办法之一是引进代数结构以提高建模的效率，相应的模型有匹特里网、有限递归过程、通信序贯进程等. 特别是匹特里网，可以加入时间、随机变量等因素以增强其描述能力，又可提高其结构化程度以减少节点数，使模型更加精炼，因此特别得到重视. 对于难以解析或计算求解的问题，计算机仿真通常是惟一的研究方法，已经开发了多种专用软件. 由于这类仿真数据处理量大，耗费机时，又有扰动分析等方法直接用于优化分析. 总之，离散事件动态系统的问题是多方面、多层次的，相应的模型和方法也是多种多样的.

离散事件动态系统的若干有关概念虽早已有之，但作为一个统一的学科加以命名，是1980年哈佛大学华裔教授何毓琦提出的. 它的实际背景包括生产制造系统、计算机和通信网络、电子化防务指挥系统、公共设施服务系统以及复杂的多级控制系统、柔性制造系统等各种高新技术问题. 对这些复杂系统，不仅要求解决从逻辑特性设计、软件实现、校验，直到随机品质分析、优化等各层次的问题，而且常会遇到常规的连续变量动态系统和离散系统相互耦合的问题. 因此，各层次模型和数学工具的集成以及混杂系统的研究，已经引起了广泛的重视.

离散事件动态系统逻辑层次模型(logical level model for discrete event dynamic system)　描述离散事件动态系统的逻辑演化和控制的各种数学模型. 这种模型应包含以下几种要素：

1. 状态集合. 常为离散的有限或可数集合，其元素间一般不必定义任何结构和运算.

2. 事件集合. 另一个离散集合,其元素事件按某些自发或预设的机制异步地出现并可能触发状态的演变.

3. 演化规则. 即由事件出现导致系统原状态转移为新状态的规则,这种规则可能是确定性的或随机性的,其结果将形成若干状态串或轨线.

4. 控制目标. 通常可用达到某特定状态子集(通常称标识状态集)或其状态轨线属于(或不属于)某预定轨道(即状态串)的集合来加以描述.

5. 控制机制. 即影响事件出现或状态演化过程的各种机制.

通常由人为设计,以满足预定的控制目标的确定性模型,包括有限(状态)自动机、描述并发过程的匹特里网,以及它们的各种扩充. 又如有限递归过程、通信序贯进程、通信系统代数等,几乎原用于理论计算机科学的逻辑工具和成果均可借用. 此外,常见的马尔可夫链也可作为一种随机的逻辑演化过程的模型.

离散事件动态系统逻辑监控理论(logical supervisory control theory for discrete event dynamic system) 离散事件动态系统逻辑层控制的一种理论. 它以离散的事件或状态过程满足预定逻辑规范为目标而进行控制器设计. 根据被控对象的不同描述模型(参见"离散事件动态系统逻辑层次模型")和对系统观测信息及控制途径的不同,逻辑监控的方式也有多种.

对离散事件动态系统的有限自动机模型,一种监控途径是假定所有可能发生的"事件"中,一部分是无法加以防止的,称为不可控事件. 其余则是可以取消和防止的,称为可控事件. 而可用的信息包括系统输出的事件符号串信息,或系统状态演变的轨迹信息. 这时逻辑监控问题归结为如何由可得到的信息来决定是否要对某些可控事件加以阻止或取消. 控制目标可表示为由此生成的事件串集合(称为子语言)分别不能小于及不能大于某些给定的集合. 控制机制可以是用专门设计的另一称为监控器的有限自动机来实现的动态逻辑反馈,也可是按系统状态确定的逻辑反馈. 这种监控理论的主要结果,包括用所谓可控子语言概念表达的监控器存在性的充分必要条件;逻辑监控和状态反馈控制间的关系;等等. 由于这类问题的主要困难在于计算复杂度过高,又导致对其进行分解、合成的分散监控或递阶协调监控的研究. 该方法曾被用于一种复杂的半导体生产设备的逻辑设计,根据问题特点,该设备中引入一种可强制使能的事件,简化了控制器结构.

用其他各种逻辑层次模型也可提出相应的逻辑控制问题. 例如,常用的匹特里网(参见"匹特里网"),具有更强的描述能力和分析及综合方法,并已形成一套递阶建模和控制综合的方法和工具.

匹特里网(Petri net) 一种逻辑模型. 它适于描述和分析异步并发系统. 自 1962 年德国数学家匹特里(Petri,C. A.)在他的博士论文"用自动机通信"中首次提出这种模型以来,已被广泛地用于计算机通信和自动控制等许多学科领域. 特别是当今迅速发展的高速计算和高速通信对并行技术的迫切需求,使得并发模型——匹特里网越来越受到广泛重视,成为并行技术十分重要的理论模型. 匹特里网从结构上说,实际上是一种有向二部图,它包含两类结点,分别称为位置结点和变迁结点. 用图形表示时,位置结点表示为小圆圈,而变迁结点表示为小矩形或短横,位置结点与变迁结点之间的关系用有向边表示. 以上只描述了匹特里网的结构,为了描述其动态行为,在网中还要引入一种称为托肯的小黑点,它存在于网中某些位置结点. 托肯在位置集合中的初始分布称为初始标识(初态). 托肯可按以下规则在网中流动,从而表现了系统的动态信息流特征:

1. 网中某个变迁只有当它的所有输入位置(位置结点到变迁结点间存在有向边)都含有足够托肯时才是可引发的.

2. 一旦该变迁引发,其每个输入位置中的托肯流出,而变迁的每个输出位置将流入托肯. 图 1 给出了一个具有初始标识的匹特里网,其中变迁 t_3 是可引发的,t_3 引发之后的匹特里网及其标识如图 2 所示.

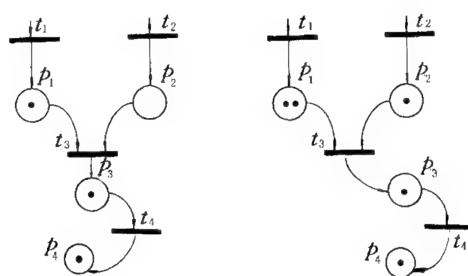


图 1　　图 2

为了方便应用,对基本匹特里网模型还可以进行一些扩充. 例如,对位置引入容量函数,对有向边引入权函数,对变迁引发引入时间因素以及随机因素等. 总之,匹特里网已经形成比较完整的体系,有丰富的支持理论和分析方法,便于应用. 总的来说,匹特里网适于作为由事件驱动、状态演变的离散并发系统描述和分析的模型,往往以位置表示条件,变迁表示事件,位置中有托肯(或称令牌)表示条件成立. 引发规则实质上是一种推理机制,其特点概括起来有:

1. 匹特里网是以图形表示的模型,直观性强,目前已有的图形软件可用于匹特里网的仿真和分析.

2. 匹特里网中的托肯流动表现了系统的动态演

变过程.

3. 匹特里网能准确地刻画系统的一些重要特性,如并发、冲突、同步、异步、死锁、饥饿和溢出等.

4. 有一套严格的数学理论和分析方法,支持对系统模型的各种性质的分析和性能评价.

匹特里网的主要性质包括可达性、有界性和安全性、活性、公平性等,而它的主要扩充形式有时延匹特里网、时间匹特里网、随机匹特里网、谓词变迁网、颜色匹特里网、高级网等.

匹特里网的可达性(reachability of Petri net) 匹特里网的基本性质之一. 即若 M_0 是它的一个状态(即标识),由 M_0 通过引发变迁 t 演变到状态 M'_1,则称状态 M'_1 是由 M_0 一步可达的. 若存在变迁序列 $\sigma=t_1,t_2,\cdots,t_k$,使得 M_0 引发变迁序列 σ 到达状态 M'_k,则称 M'_k 是从 M_0 可达的. 若 M_0 是匹特里网的初始状态,则从 M_0 可达的所有状态(包括 M_0)的集合称为 M_0 的可达状态集,并记为 $R(M_0)$. 匹特里网的可达性反映了匹特里网是否能实现某种功能,是其最重要的特性之一.

匹特里网的有界性(boundedness of Petri net) 匹特里网的基本性质之一. 若存在正整数 k,使其在任何可达状态下,任何位置的托肯数都不大于 k,则称匹特里网是有界的. 特别当 k 为正整数 1 时,则称匹特里网是安全的. 匹特里网的有界性(安全性)反映了被模拟系统运行过程中对有关资源的容量要求.

匹特里网的安全性(safeness of Petri net) 见"有界性".

匹特里网的活性(liveness of Petri net) 匹特里网的基本性质之一. 若对其中任何变迁 t,任何可达状态 $M\in R(M_0)$,都存在一个从 M 可达的状态 M^1,使得 t 在 M^1 下是可引发的,则称该匹特里网是活的. 匹特里网的活性对应着系统的无死锁性,即不会出现死状态或局部死循环. 匹特里网的活性是其重要的动态性质,往往也是系统的基本要求.

匹特里网的公平性(fairness of Petri net) 匹特里网的基本性质之一. 即如果存在一个正整数 k,使得对任何可达的状态 $M\in R(M_0)$,及其任何可引发的变迁序列 $\sigma\in T^*$,都有性质:当任何一个变迁 $t_i\in T$ 在 σ 中出现的次数都是 0 时,则任何一个其他变迁 $t_j\in T$ 在 σ 中出现的次数不超过 k,这时,称该匹特里网是公平的. 匹特里网的公平性反映了系统的无饥饿性. 如果一个资源共享系统的模型是一个公平匹特里网,那么该系统各分支在共享资源的竞争中就不会出现饥饿现象.

计时事件图(timed event graph) 一类常用的匹特里网. 其形式定义为:

1. 每个位置 p 只有一个前置变迁和后续变迁.

2. 各弧权均为 1.

3. 位置容量不限.

4. 每个位置 p 具有(非负整数)时延 t,令牌在 p 中产生后,需经 t 时后才可使用.

5. 启动后每个初始令牌可随时使用.

6. 变迁不占用时间.

这种匹特里网没有竞争和选择问题,并可方便地用双子代数来描述和研究(参见"双子代数").

赋时匹特里网(timed Petri net) 一类特殊的匹特里网. 由拉姆钱达尼(Ramchandani, C.)于 1974 年提出. 模型中对每个变迁都赋予有限的引发时延值,从而它可用于评估所描述过程的运行时间,并对其性能进行分析、优化.

随机匹特里网(stochastic Petri net) 一类特殊的匹特里网. 指具有随机演化规则的匹特里网,由莫洛伊(Molloy, M. K.)于 1982 年提出. 模型中对每个变迁的引发都规定了一个随机的时延参量. 大多数随机匹特里网模型的性能分析方法均基于其状态空间与马尔可夫链的同构性质,为此,通常需要求相关的连续时间随机变量是指数分布.

时间匹特里网(time Petri net) 一类特殊的匹特里网. 即描述变迁的时间特性的匹特里网. 由梅林(Merlin, P.)等人于 1976 年提出. 该模型规定每个变迁都对应着一个时间区间$[a,b]$,任何一个变迁,当它使能之后,它在时间区间$[a,b]$内便具有连续使能权. 使用这种网,梅林等讨论了一些计算机系统和进程中的可达性问题. 应该说,时间匹特里网又是时延匹特里网的一种推广.

颜色匹特里网(color Petri net) 一类特殊的匹特里网. 即具有以颜色区分的多种托肯的匹特里网,由詹森(Jensen, K.)于 1981 年提出. 它在原有基网的基础上赋予以下新内容:

1. 定义色函数 C: $P\cup T\rightarrow$颜色集的幂集合,使得对所有 $t\in T$,$C(t)$为 t 上所有可能的出现颜色之集合;对所有 $p\in P$,$C(p)$为位置 p 中所有可能的托肯颜色之集合.

2. 同一位置可包含多个不同颜色托肯的匹特里网,匹特里网中的托肯根据其作用分类,同类的托肯用同一颜色表示.

3. 每个变迁上标明所有可能的转换所改变的托肯的颜色及其个数.

4. 根据变迁上的规定决定在变迁发生时各位置失去或得到的托肯的颜色及数量.

5. M_0 是其初始标识(状态),使得 $\forall p\in P$,$M_0(p)$是 p 的托肯匹特里网的色集合上的多重集.

高级网(high-level Petri net) 一类复杂的匹特里网. 它具有更强描述能力和变迁过程的各种扩展的定义,具有更丰富的内涵,并降低网结构的复杂

度. 前述谓词/变迁网、颜色匹特里网、随机匹特里网和时间匹特里(延)网等均属高级网之列. 这些网更贴近于实际问题,可降低模型的复杂度.

谓词/变迁网(predicate/transition net)　一类高级网. 它的网结构(称为基网)满足如下附加要求:

1. 每个位置对应着一个可变谓词,其中的托肯代表使谓词成真的个体. 个体各不相同,亦即任何两个托肯都不相同.

2. 每个变迁对应着有固定外延的静态谓词. 这种谓词中出现的自由变量可用个体来替换,每个可行替换对应着该变迁的一次引发.

3. 其弧上可以出现自由变量,它们的值由相应变迁上的可行替换决定.

4. 其初始标识(状态)M_0 可使得任意 n 元谓词 $p \in P$, $M_0(p)$是 n 元符号和.

通信顺序进程(communication sequential process)　简称 CSP. 一类逻辑过程模型. 是由英国科学家赫尔(Hoare, C. A. R.)于 1978 年提出的一种命令式语言. 一种 CSP 程序就是一个进程,每个进程可以平行地分解为许多子进程,子进程之间以进程运算符相联系. 子进程又可进一步分解为更深一层的子进程,这种进程嵌套可达到任一深度. CSP 的最大特点是有一对通信原语,沟通进程之间的联系. CSP 还提供了一组比较丰富的不确定和平行运算操作,其中包括不确定选择符、变进程运算符、或进程运算符、选择进程运算符、并发进程运算符. 为了增强描述能力,CSP 还有一些其他设备,如递归定义的进程、限制进程、屏蔽进程、换名进程等.

通信演算系统(communication calculus system)　简称 CCS. 一类逻辑模型. 由美国科学家米尔纳(Milner, R.)于 1980 年提出的一种函数式语言. 它的基本成分是项,或称动程,含自由变量的动程称动程表达式. 动程的组合仍是动程,其组合深度可以是任意的. CCS 也有一对通信原语,它由两组标号组成. 其中一组是名字的集合,常表示为 $\Delta=\{\alpha,\beta,\gamma,\cdots\}$;另一组是加上横线的名字集合,常表示为 $\bar{\Delta}=\{\bar{\alpha},\bar{\beta},\bar{\gamma},\cdots\}$, Δ 称为输入港口集, $\bar{\Delta}$ 称为输出港口集,它们是互补的. 动程间通信时并不指定对方动程的名字,只指定对方港口的名字,只有相同名字的输入和输出(不考虑上横线)之间才可以通信. 相对说来,CCS 的复合操作比较少,只有一个并行操作,相当于 CSP 的并发操作以及一个选择操作,相当于 CSP 的或进程运算,但这并不意味 CCS 的描述能力不如 CSP 强.

时段演算(duration calculus)　一种实时区间时态逻辑. 由中国数学家周巢尘和英国科学家赫尔(Hoare, C. A. R.)等人于 1991 年提出的. 它将布尔函数在区间上的积分进行形式化,从而可用来描述和推导离散状态系统的实时和逻辑特性. 具体地,时段演算将积分当做从状态函数和区间到实数的函数,其形式定义为

$$\int : S \to (I \to \mathrm{R}),$$

其中 S 表示状态集(为布尔函数), I 表示闭区间集. 因此,区间时态逻辑被采用作为基础逻辑,并且区间函数

$$\int S, \int P, \cdots$$

变成了演算的区间变量,这里 S 和 P 是状态,此外还要假设状态具有有穷变化性以保证可积性条件.

时段演算具有六个公理:

1. $\int 0 = 0.$

2. $\int P \geqslant 0.$

3. $\int P + \int Q = \int (P \vee Q) + \int (P \wedge Q).$

4. $\left(\int P = r + s\right) \Longleftrightarrow \left(\int P = r\right), \left(\int P = s\right)$
$(r, s \geqslant 0).$

5. $(\lceil\ \ \rceil \vee \lceil P \rceil, \mathrm{true} \vee \lceil \neg P \rceil, \mathrm{true}).$

6. $(\lceil\ \ \rceil \vee \mathrm{true}, \lceil P \rceil \vee \mathrm{true}, \lceil \neg P \rceil),$

其中 $\lceil\ \ \rceil \triangleq (l = 0)$.

时段演算还有以下几种扩展的形式:

1. 扩展时段演算. 对时段演算在如下几个方面进行了推广:

1) 将$\rceil \cdot \lceil$从布尔函数推广到相等、不等、连续等性质. $\rceil H \lceil$在一个区间上成立,当且仅当区间不是点区间,并且性质 H 在区间内处处成立.

2) 为了描述微分方程的边界条件,引入区间变量 $b.v$(和 $e.v$),用于表示函数 v 在非点区间上的开始点(结束点)和右极限(左极限). 它满足如下公理和规则:

① $\rceil \mathrm{continuous}(f) \lceil \wedge (l>0 \wedge e.f=r_1),$
$(b.f=r_2 \wedge l>0) \Rightarrow r_1=r_2.$

② $R(\rceil H(f) \lceil, b.f, e.f, l)$是不带切变的公式.

③ 为了从微分方程导出状态的积分,对任意状态 P 引入一个函数 f_P,其定义如下:

$$\rceil\ \ \lceil \vee \rceil \mathrm{continuous}\ (f_P) \lceil) \wedge \Box (\rceil P \lceil$$
$$\Rightarrow (f_P=1)) \wedge \Box (\rceil \neg P \lceil \Rightarrow (f_P=0)),$$

从而定义

$$\int P \triangleq e.f_P - b.f_P,$$

称这种用于刻画分段连续性和可微性的演算为扩展的时段演算.

2. 平均值演算. 以布尔函数 F 的平均值(记为 $\overline{F}$)代替原来演算在区间上 F 的积分 $\int F$,从而扩展成带点值的时段演算,称其为平均值演算.

在区间上,积分总是蜕化成 0,而一个函数在点区间上的平均值却代表着函数在那一点上的函数值. 函数 F 在区间$[a,b]$上的平均值定义如下:

$$\overline{F}[a,b]=\begin{cases}\int_a^b F(t)\mathrm{d}t\Big/(b-a) & (b>0),\\ F(a) & (b=a).\end{cases}$$

函数 F 的积分$\int F$ 可以用 $\overline{F}$ 和 l 来定义

$$\int F \triangleq \overline{F} * l.$$

用 $\overline{F} * l$ 替代时段演算公理中的$\int F$,再引入如下定义点值的公理,便构成了平均值演算的一个相对完全的公理系统:

$\rceil F\lceil^0 \triangleq \rceil\ \lceil \wedge (\overline{F}=1)$;

$\rceil F\lceil \triangleq (l>0) \wedge \neg((l>0),(\rceil\ \lceil \wedge \overline{F}=0),(l>0))$;

$\rceil F\lceil^* \triangleq \rceil F\lceil^0 \vee \rceil F\lceil$;

$\rceil\ \lceil \Rightarrow (\rceil F\lceil^0 \vee \rceil\neg F\lceil^0)$;

$\rceil F\lceil^* \Rightarrow (\overline{F}=1)$.

3. 概率时段演算. 它为设计人员提供了一套规则,用来推导和计算一个不可靠系统满足时段演算所表示的需求的概率. 这里,不可靠系统是用概率自动机来描述的.

袋(bag) 集合概念的一种推广. 同集合一样,袋是某论域中一些元素的总体,但不同的是在袋中同一元素可以重复出现多次. 袋的概念由策尔夫(Cerf,V.)等人于 1971 年提出,后来被皮特逊(Peterson,J. L.)用于匹特里网中引发序列的无序表达等问题. 若以$\#(x,B)$表示元素 x 在袋 B 中出现的次数,则当$\forall x\in B$:$0\leqslant \#(x,B)\leqslant 1$ 时,B 就蜕化为普通的集合. 袋的基本概念可定义如下:

1. 属于关系.

当$\#(x,B)>0$ 时,称 x 属于 B,记为 $x\in B$;

当$\#(x,B)=0$ 时,称 x 不属于 B,记为 $x\notin B$.

2. 基数.

$$|B|=\sum_x \#(x,B).$$

3. 包含. $A\subseteq B$ 当且仅当对任意 x,

$$\#(x,A)\leqslant \#(x,B).$$

4. 相等. $A=B$ 当且仅当 $A\subseteq B$ 且 $B\subseteq A$.

5. 袋的运算又包括:

1) 并 $A\cup B$:

$\#(x,A\cup B)=\max(\#(x,A),\#(x,B))$.

2) 交 $A\cap B$:

$\#(x,A\cap B)=\min(\#(x,A),\#(x,B))$.

3) 和 $A+B$:

$\#(x,A+B)=\#(x,A)+\#(x,B)$.

4) 差 $A-B$:

$\#(x,A-B)=\#(x,A)-\#(x,B)$.

混杂动态系统(hybrid dynamic system) 简称 HDS. 一类兼有离散事件和连续变量两种运行机制的动态系统. 由于复杂系统中通常存在不同层次、不同类型的各种子系统,需要用定性定量相结合的方法来处理. 定性行为常用符号、逻辑、模糊等离散型变量来描述,而定量行为则可用连续变量及相关的微分或差分方程来描述. 从而混杂动态系统是研究复杂系统的一种重要的数学模型和理论方法. 混杂动态系统通常可分为两类:第一类中,离散事件和连续变量两类子系统大体可以分开,其间的相互作用可通过一些离散-连续转换接口来描述,这种系统称为松耦合的,常见于逻辑控制、模糊控制或粗粒度量化观测的工程系统中;第二类则为紧耦合混杂动态系统,其中离散、连续两种变量可能有统一的表达方式,并可进行混合运算处理,对此已设计有多种专门的仿真语言. 总的说来,连续变量运动过程大多反映被控物理过程的机制和行为,而离散部分则反映人们按一定要求而设计的信息采集、逻辑判断和施行控制的机制.

混杂动态系统的基本问题可概括为:对给定的连续变量被控对象及反映设计要求的品质规范和各种物理上或逻辑上的约束,求出一种用数字或逻辑开关程序实现的控制算法,以保证系统满足预定的规范要求. 从历史上说,虽然这类系统早已存在于可编程逻辑控制、计算机控制等实际问题中,但并未有系统化的完整理论. 离散事件动态系统理论的发展为离散过程提供了一个比较统一的理论框架,人们已自然地提出研究混杂动态系统的问题. 目前已有的模型不下几十种,大体可分为以下几类:

1. 自动机和状态迁移系统.

2. 动态系统.

3. 代数结构.

4. 程序语言.

进一步,由物理过程采集信息的机制、连续变化部分的动力学性质、连续部分或离散部分是否有导致轨道不惟一的不确定因素、控制作用是连续量、离散量还是组合式、对系统行为的要求规范是连续型、离散型还是混合型,进而提出各种不同的具体模型和问题. 混杂动态系统的实际背景包括连续流程型工业的控制管理一体化系统、磁盘读写头伺服和解码系统、交通和通信系统、变结构控制系统等.

极大代数(max-algebra) 一类描述离散变迁时间特性的代数系统. 记为 $D=\{\overline{R},\oplus,\otimes\}$,它定义

于 $\overline{R}=\mathrm{R}\cup\{\varepsilon\}$，其中 R 为实数集，$\varepsilon\triangleq-\infty$，并对 $a,b\in\overline{R}$ 分别用 $a\oplus b\triangleq\max\{a,b\}$ 和 $a\otimes b\triangleq a+b$ 来定义其加法和乘法. 这两种运算满足结合律、交换律及乘法对加法的分配律，ε 为加法零元，$e\triangleq 0$ 为乘法么元，从而对之可以建立常规的算术运算系统. 由于其加法具幂等性 $a\oplus a=a$，而且没有逆运算减法，极大代数又有许多独特的性质. 事实上，它是介于格和线性代数之间的一种代数系统.

由于实际问题中经常遇到求"最长路径"、决定某事件发生时刻的若干个前提事件的"最迟发生时刻"等运算，均可方便地用极大代数表示，从而它已成为相应问题有力的描述和分析工具. 它的基本思想自 1954 年开始即有人提出，而后由英国康宁汉-格林(Cuninghame-Green，R.)建立了完整的理论体系.

在上述运算基础上，可类似地定义极大代数上的矩阵运算和线性多变量系统的形式描述，并研究相应的矩阵本征值问题，系统的周期行为、稳定性、反馈控制、实现理论等基本问题. 这些问题在各种离散事件过程、时间层次模型的分析、设计、控制中有重要应用.

与极大代数类似，可定义极小代数、极大-极小代数以及更一般的双子代数(参见相关条目). 相应的 $\overline{R}$ 可改为 $\underline{R}=\mathrm{R}\cup\{+\infty\}$，或者整数集的扩充 $\overline{Z}=\mathrm{Z}\cup\{-\infty\}$或 $\underline{Z}=\mathrm{Z}\cup\{+\infty\}$.

双子代数(dioid algebra)　极大代数的推广. 常记为代数系统 $\mathscr{D}=(D,\oplus,\otimes)$. 在集合 D 上定义的运算"加法"($\oplus$)和"乘法"($\otimes$)满足：

1. 结合律.
2. 加法交换律.
3. 乘法对加法的左、右分配律.
4. 存在零元 ε，对$\forall a\in D$ 有
$$a\oplus\varepsilon=a,\ a\otimes\varepsilon=\varepsilon\otimes a=\varepsilon.$$
5. 存在么元 e，有 $e\otimes a=a\otimes e=a$.
6. 加法具幂等性 $a\oplus a=a$.

进而称双子为完备的，若其对任意无限多元的"和"封闭且乘法对此无限和具分配律. 又称双子为阿基米德的，若$\forall a,b\in D$，均有 $c,d\in D$，使有 $a\otimes c\geqslant b$，$d\otimes a\geqslant b$，这里$\geqslant$号意为 $x\geqslant y\Leftrightarrow x\oplus y=x$，这时亦称 $y\triangleq x\wedge y$ 为 x,y 之交. 交与和是对偶运算. 还可称双子为分配的，若其"和"对"无限交"及"交"对"无限和"的分配律均成立. 基于上述运算可以建立双子代数上的矩阵和线性系统模型，并研究其"线性"代数性质和动态行为反馈控制等问题.

双子代数的系统研究和应用始于康宁汉-格林(Cuninghame-Green，R.)和科恩(Cohen，G.)等人. 其具体形式有多种，较重要者如极大(小)代数、2D 域$\ll\gamma,\delta\gg$代数等. 它们适用于描述和分析具确定性时间的离散事件过程，诸如柔性制造系统、计划调度系统等.

极小极大 $\ll\gamma,\delta\gg$ 代数(min-max $\ll\gamma,\delta\gg$ algebra)　一种特殊的双子代数，记为
$$\mathscr{M}=(M,\oplus,\otimes),$$
其中 $M=\{\bigoplus_i\gamma^{n_i}\delta^{t_i}\mid n_i\in\underline{Z},\ t_i\in\overline{Z}\}$，$\overline{Z}=\mathrm{Z}\cup\{-\infty\}$，$\underline{Z}=\mathrm{Z}\cup\{+\infty\}$. "加法"运算定义为将所有 $\gamma^{n_i}\delta^{t_i}$ 型的同类项"加"起来，并按下列规则进行归并化简：
$$\gamma^n\delta^t\oplus\gamma^n\delta^{t'}=\gamma^n\delta^{\max(t,t')},$$
$$\gamma^n\delta^t\oplus\gamma^{n'}\delta^t=\gamma^{\min(n,n')}\delta^t.$$
当 $n>n'$ 且 $t>t'$ 时，$\gamma^n\delta^t$ 和 $\gamma^{n'}\delta^{t'}$ 是不可比较、亦无法归并的，它们将同时保存下来. 乘法运算则定义为按规则 $\gamma^n\delta^t\otimes\gamma^{n'}\delta^{t'}=\gamma^{n+n'}\delta^{t+t'}$ 展开乘积，然后再用上述"加法"规则进行归并化简.

可以验证，上述定义的代数系统符合双子定义的各条件. 还可证明，这个双子是可交换的、完备的、分配的和阿基米德的.

这种定义的实际含义可解释为信息元 $\gamma^n\delta^t$ 对应 nt 平面上一个点，它表示事件发生之计次为 n 时的最短时间为 t，或对偶地，计时为 t 时事件最多可能发生的次数为 n. 因此信息元之间有些可比较(从而可归并)，有些则不可比较，只好全部保留.

同样可建立极小极大$\ll\gamma,\delta\gg$代数的矩阵运算、线性系统模型及相关的动力学和控制理论.

极大代数矩阵本征值问题(eigenvalue problem of matrix in max-algebra)　由极大代数导出的一类矩阵本征值问题. 按照极大代数中的加法$\oplus$和乘法$\otimes$的规则，可以和常规线性代数类同地定义矩阵及其运算. 例如，若 $A=(a_{ij})_{m\times p}$，$B=(b_{ij})_{p\times n}$，则
$$A\otimes B\triangleq(c_{ij})_{m\times n},$$
其中
$$c_{ij}=\bigoplus_{k=1}^{p}(a_{ik}\otimes b_{kj}).$$
许多实际问题可以归结为研究由下列矩阵关系定义的线性变换：$x(t+1)=A\otimes x(t)$，其中 $x(t)\in\overline{R}^n$ 可理解为第 t 拍时的 n 维状态向量，A 为 $n\times n$ 矩阵.

自然可提出该矩阵 A 的本征值问题：能否找到实数 λ 和某向量 x，使得 $A\otimes x=\lambda\otimes x$? 若能找到且令$x(0)=x$，则有
$$\begin{aligned}x(t+1)&=A\otimes x(t)=\lambda\otimes x(t)\\&=\lambda^2\otimes x(t-1)=\cdots\\&=\lambda^{t+1}\otimes x(0).\end{aligned}$$
这里极大代数意义下的 $\lambda^k\triangleq\lambda\otimes\lambda\otimes\cdots\otimes\lambda=k\lambda$，表明每演化一拍，$x$ 的各分量均增加相同的值 λ. 由于极大代数描述的问题中，$x(t)$常表示第 t 拍时各事件发生的时刻，若求出本征值和本征向量，则可断言对应的系统行为进入了一种以 λ 为周期的周期态，而

这通常是人们期望并常在实际中观察到的. 当系统能进入某种周期态或周期态的组合时,则称此(极大代数意义下的)系统为稳定的.

极大代数矩阵本征值问题与普通线性代数有完全不同的结论. 为叙述这些结果,首先要将矩阵 A 与下列加权有向图对应起来. 该图有 n 个结点,分别代表 x 的一个分量. 仅当矩阵 A 的 (i,j) 元素 $a_{ij}\neq-\infty$ 时,图中有一条由结点 i 到结点 j 的权为 a_{ij} 的有向边. 对该图的每一条长为 l 的回路 $(i_0,i_1,\cdots,i_l=i_0)$, 定义

$$\lambda=\frac{1}{l}\sum_{j=1}^{l}a_{i_{j-1}i_j}$$

为该回路的权(其中运算为普通算术意义下的).

可以证明,当该图为强连通亦即矩阵 A 为不可约时,各回路最大的权 λ 即为该矩阵的惟一本征值. 按定义它可简洁地表达为

$$\lambda=\sum_{n}{}_{\oplus}(\operatorname{tr}A^n)^{1/n},$$

其中所有运算都是在极大代数意义下的.

当该图不是强连通时,其本征值不仅应为某回路 r_i 的平均权重 λ_i,而且这些 r_i 到其他平均权重大于 λ_i 的回路均无通道. 反之,这些条件也保证了 λ_i 必为本征值. 应当指出,本征向量的求法也是比较复杂的. 对这种极大代数意义下的"线性"系统,亦可用状态反馈或输出反馈来使受控系统稳定并具有指定的本征值(运行周期).

星运算(star operation) 双子代数特别是极大代数中的一种重要运算. 双子代数中元素 a 的星运算 a^* 定义为

$$a^*\triangleq e\oplus a\oplus a^2\oplus\cdots.$$

可以证明

$$(a^*)^*=a^*;\ \varepsilon^*=e;\ (a^*)^k=a^*,\ \forall k\in Z^+,$$

以及令 $a^+\triangleq aa^*$ 时 $a^*=e\oplus a^+$, $a^*\geqslant a^+$ 等. 对矩阵双子亦可类似定义. 星运算主要用于解双子中的线性方程,如矩阵方程 $x=Ax\oplus b$ 的解为 $x=A^*b$,但计算 A^* 并不简单,需要特别的技术.

随机赋时状态自动机(stochastic timed state automaton) 状态自动机概念的一种扩充. 其定义为六元组 $(\mathscr{E},\mathscr{X},\Gamma,p,p_0,G)$,其中 $\mathscr{E}$ 为可数的事件集;$\mathscr{X}$ 为可数的状态空间;$\Gamma(x)\subseteq\mathscr{E}$ 为对 $\forall x\in\mathscr{X}$ 定义的可行或使能的事件集;$p(x';x,e')$ 为当事件为 e' 时由 x 到 x' 的状态转移概率,显然当 $e'\notin\Gamma(x)$ 时 $p(x';x,e')=0$;$p_0(x)$ 为初始状态 x_0 的概率分布;$G=\{G_i:\ i\in\varepsilon\}$ 为随机时钟机构,它表示事件 i 由第 $k-1$ 次发生到第 k 次发生的间隔 $v_{i,k}$ 的概率分布.

这种自动机的演化过程为:由任一状态 x 出发,按 $\Gamma(x)$ 中由 G 随机决定的最早发生的事件 e' 触发,并由概率 $p(x';x,e')$ 决定转移到新状态 x'. 与此同时,随机时钟机构和新的 $\Gamma(x')$ 都要相应地加以更新.

随机离散事件动态系统(stochastic discrete event dynamic system) 简称随机 DEDS. 按随机规律进行状态迁移的离散事件系统. 诸如事件发生的种类和时刻服从随机分布、状态迁移映射具随机性等. 在实际问题中,由于顾客或工件的到达、加工服务时间、机器的工件或修复等往往都带随机性,随机 DEDS 的研究具很大的现实意义.

随机 DEDS 的基本模型是排队网络(参见"排队网络"). 它由顾客来流分布特性、排队规则、队长容量、服务时间分布、服务台数量及路径等来刻画. 当其中各概率分布均为负指数,从而具有无后效性时,它又可用马尔可夫链描述,系统的状态则用各服务台前的顾客数(向量)来表示. 排队论已得到许多有关系统中顾客数、顾客平均等待或停留时间、平均队长等品质指标. 对一般情形,由于可能的状态数很大,其联合概率分布的形式十分复杂,计算极为困难. 实际上,较容易操作的只是具有乘积形式解的排队网络. 其中,联合分布蜕化为各服务台队长分布的乘积,从而较易计算. 常见的这类网络包括开杰克森网、闭杰克森-戈登-诺威尔网、BCMP 网等. 对乘积解已有一些快速算法,如卷积算法、均值算法等.

除了上述分析结果,近来还研究了许多有关系统设计、品质评估和优化、调度控制等问题的方法. 由于问题都较复杂,实际中常采用随机仿真试验来辅助求解. 特别在系统设计时,经常要回答"如果…将…"(what if)问题. 例如,在参数优化时为了测量品质指标对参数的灵敏度,需要对各种参数值进行多次试验并作比较,这又涉及产生随机样本等大量计算的要求. 对此已提出了几种由单个样本轨道估计灵敏度的巧妙方法,诸如扰动分析、似然比方法等.

对仿真执行过程的形式描述,导致一种较一般的广义半马尔可夫过程模型. 它事实上包括事件-状态相互作用机制的界定(称为广义马尔可夫结构)和各种随机因素的描述这样两大部分. 它体现了离散事件系统各层次模型间的关系,也可认为是离散事件系统一种较为完整的描述. 此外,若干特定结构的随机离散事件动态系统可以在相应的自动机、匹特里网或极大代数模型结构的基础上加入各种随机因素来加以描述,从而形成随机匹特里网、随机赋时自动机、极大代数上的随机线性系统等. 它们基本上又都能纳入广义半马尔可夫过程的仿真模型.

排队网络(queueing network) 随机离散事件动态系统模型的一种. 其主要特点为各事件发生时间服从已知的概率分布. 这种系统可以用不同种类

顾客沿各自的路径到达不同的服务台接受服务这种形象化的过程予以直观理解．其形式化描述需阐明以下几方面特征：

1．顾客（亦泛指待加工零件等）的种类、到来数量的特征、优先等级和规则等．

2．服务台（亦指加工机器等）的数量和服务时间的统计分布．

3．排队规则，如是否先到先服务、顾客是否因等待和服务时间过长而中途离去等．

4．排队空间及系统中总顾客数的限制．

5．在多服务台形成服务网络情形还有关于顾客接受多种服务的路径和规则等，研究的课题包括系统性能指标，如平均排队长度和顾客保留时间的统计分析、系统参数和结构设计以及调度优化的规则等．

对排队网络的描述有一套符号体系．例如，$M/G/1$ 表示顾客来到的间隔时间为负指数分布（M）、服务时间为一般分布（G）、单个服务台（1）等．排队网络问题早已由电话服务系统等问题提出和发展，但难度甚大，连看来简单的 $G/G/1$ 问题至今未有通用的解法．排队网络的状态一般用各服务台前各类顾客排队长度的联合分布来描述．只有当该联合分布可以分解为各个队列长度分布的乘积形式，即各队长分布相互独立时，问题才较容易解决．因此，研究具乘积形式解的排队网络类型是一个重要的理论和实践的问题．例如，杰克森网络和 BCMP 网络（参见“BCMP 网络”）都是著名的具有乘积形式解的例子．

排队网络模型在生产调度、计算和通信系统优化等方面有重要应用．随机仿真试验仍是一种主要研究途径，最近发展了一些基于仿真的统计优化技术，可望达到实用的目的（参见“广义半马尔可夫过程”、“扰动分析”）．

BCMP 网络（BCMP network）　一类排队网络．它以首先提出的四位合作者巴斯科特（Baskett, F.）、强迪（Chandy, K. M.）、蒙兹（Muntz, R. R.）和巴拉修斯（Palacios, F. G.）的名字首字母而命名．这是一种具 K 类不同顾客的闭网络，每类顾客各具自己的路径概率和在各结点的服务率．其中包含的结点可能有四类：

1．具指数分布服务时间、服从先来先服务规则的单服务台．

2．具任意可微的服务时间分布、处理器共享排队规则的单服务台．

3．在第 2 类中将排队规则改为允许抢先和事后继续完成的后来先服务型．

4．具无穷多服务台及任何可微的服务时间分布的结点．

BCMP 网络给出了一大类具乘积形式解的排队网络，从而引起了人们的注意．应指出，这类网络不一定具有马尔可夫性质．

广义半马尔可夫过程（generalized semi-Markovian process）　简称 GSMP．马尔可夫过程的一种推广．它是一种具状态空间 $\mathscr{X}$，并由随机赋时的状态自动机（$\mathscr{E}, \mathscr{X}, \Gamma, p, p_0, G$）（参见“随机赋时状态自动机”）生成的随机过程 $\{X(t)\}$．这一概念最早由马特斯（Matthes, K.）于 1962 年提出．其后在离散事件系统研究中又得到注意，并用于样本轨道的描述和仿真．为构造样本轨道，首先应由状态空间 $\mathscr{X}$ 中的分布 p_0 经随机试验得到初始状态 x_0．这时可能触发状态转移的事件集合，记为 $\Gamma(x_0)$．对 $\Gamma(x_0)$ 中的事件按随机时钟结构 G 给出的各事件发生时间间隔（寿命）分布，随机地确定“最早”发生的事件作为触发事件 e'．按转移概率 $p(x';x,e')$ 再经随机试验得到新状态 x'．这时需要更新 $\Gamma(x')$ 及确定其寿命的时钟机构 G，并重复上述过程，如此等等．

特别地，如果每个 $\Gamma(x)$ 只含单元素（但其寿命分布可能不是指数式），则为半马尔可夫过程；若进而要求其寿命分布为指数式，则过程是马尔可夫的，这就是该术语与原有概念的关系．另外，在上述构造样本轨道过程中每步都要进行多次随机试验，即按指定分布产生随机数．这是一种相当耗费机时的工作，为此已研究了多种较为快速、方便的方法，诸如反函数法、拒斥法等．同样，在对样本轨道进行数据处理时也希望尽可能抽取更多有用信息以节省计算量，为此已经发展了扰动分析、似然比方法等．

分析 GSMP 的定义，可知当随机机制 p, p_0 和 G 均蜕化为确定性函数时，可以把事件的发生看成由某种外生的机制来决定的，则就又回到逻辑层的有限自动机模型，有人将它称为广义半马尔可夫格式．这也说明了离散事件动态系统各种模型的层次结构关系．

扰动分析（perturbation analysis）　一种数据处理和分析方法．它可用于对离散事件动态系统的仿真模型进行结构或参数优化，由美籍华裔教授何毓琦等人于 1980 年提出．通常的仿真优化方法基于在不同条件下多次随机试验结果的比较，来决定系统结构和参数优化的方向．扰动分析则通过对单一仿真轨道进行特殊数据处理，直接得到有关系统品质对参数选择的“灵敏度”的信息，从而可大幅度提高优化的效率．其要点是：分析各种扰动在系统中传播和最终在系统品质中得到实现的概率，以及用相同的随机数产生具不同特性参数的随机变量分布函数等技术．类似的思路还有鲁宾斯坦（Rubinstein, R.）的似然比方法等．

SIGNAL 语言(SIGNAL language)　可用来建立混杂系统模型的一种实时同步形式语言. 由法国的本温尼斯特(Benveniste, A.)等人于 1986 年提出. SIGNAL 语言的基本单元是信号,它们各有其信号名并由其取值序列和时钟决定,取值范围可为实数、布尔量或"缺省". 动态过程由以下基本算子描述:

1. 基本过程. 由信号间的函数与时间关系描述,诸如 $V:=Z+1$,指$\forall t$ 均有 $V_t=Z_t+1$.

2. 延迟. 用 $Z=V\ \$\ 1$ 表示$\forall t$ 均有 $Z_t=V_{t-1}$.

3. 过程组合. 例如上面两个子过程组合起来就定义了新的反馈过程($|V:=Z+1|Z:=V\ \$\ 1|$),这里竖线表示信号的连结.

4. 欠采样. 用式"$a:=b$ when$c$"表示当 c 为真时输出信号 a 取值 b,由于 a 的时钟小于 b 和 c,故称欠采样.

5. 过采样. 用"$V:=a$ default b"表示两个信号的合并,表示当 a 有值时输出 V 为 a,否则为 b. 因 V 的时钟恒大于 a 或 b,故称过采样.

6. 同步. "synchro a,V"表示 a 和 V 具有相同时钟.

在描述信号间的同步约束问题时,只考虑非布尔信号的缺省(用"-1"表示)或出现(用$+1$ 表示),以及布尔信号的值(±1),因此信号的同步演算完全在整数模 3 有限域($-1,0,1$ 或 $0,1,2$)F_3 上进行. 这里由于 $x^3=x^0$,任何函数均为阶数不超过 2 的多项式. 事实上,任何 SIGNAL 过程都可化为整数模 3 有限域 F_3 上的动态系统. SIGNAL 语言经常用于该类混杂系统的仿真,但理论分析还有一定困难.

整数模有限域上的动态系统(dynamic system on integral mod finite field)　一种有限域上的动态系统. 采用 SIGNAL 语言对混杂动态系统进行建模所得到的一组离散时间的动态方程,即

$$0=f(x_n,y_n,x_{n+1}),$$
$$0=g(x_n,y_n),$$

其中,内部状态 x_n 和系统与外界通信的信号 y_n 均随时间 n 变化,y_n 包括了输入 y'_n 和输出 y''_n. 由于各变量仅取值于某整数模 3 有限域 F_3,这里的各函数均可划归为阶数不超过 2 的多元多项式. 这些方程描述了这些变量间的"隐式"关系,为了描述系统的演化,首先要把它转化成显式表达:

$$x_{n+1}=P(x_n,y'_n,y''_n),$$
$$y''_n=Q(x_n,y'_n).$$

为此,可用 F_3 上的多项式理想理论来进行,已有不少数学结果和算法可资利用. 当 F_3 上的多项式动态系统不满足预期要求时,还可引入控制机制加以限制或改造. 对这类系统主要是通过施加禁止状态和其他约束方程来实现.

变结构控制系统(variable structure control system)　一类控制系统. 其定性结构可按一定规律变化,从而具有许多新奇性质. 这里的结构变化泛指决定系统运动形态的各种规律变化,包括控制规则或参数的突变. 最常见的是将相空间划分为具不同运动规律的区域,而当系统状态演化至区域边界时改变其运动规律. 用这种方法设计出来的控制系统,称为变结构控制系统. 变结构控制系统的独特问题之一是:由于自然原因或设计要求相轨可能迅速往返穿越切换边界,导致它在切换边界处的抖振. 物理上将沿切换边界(相空间的某个低维流形)运动,称为滑动模. 适当地设计切换规则,可以防止抖振的破坏性,并使系统具有简单的结构和良好的抗干扰性能.

切换线性系统(switch linear system)　一类混杂系统. 它由若干线性系统模型依一定条件和规则,在相空间中相互切换演化过程而构成的动态系统. 这类系统本质上是非线性的,但由于其各段轨线分属不同的线性区,有一些比较可行的分析和设计方法.

较早提出并广泛研究的变结构控制系统是切换线性系统的一种. 它按相空间区域划分来确定线性系统的不同系数. 这类系统的研究问题包括系统的稳定性、滑动模的存在、抖动和收敛性等问题,并已得到工程应用.

混杂最优控制问题(hybrid optimal control problem)　连续变量动态系统最优控制问题在混杂系统情形的推广. 按混杂系统的特点,除了常规最优控制问题的各种约束条件和费用函数外,还要计入离散开关跳变部分的品质和控制费用. 以美国密特尔(Mitter, S. K.)等人对此进行概括并提出的一种"统一模型"为例. 该模型假定连续部分的状态空间由若干流形 S_i 组成,系统在各 S_i 上的运动是连续的,并由某向量场所确定. 系统的离散部分则表现为:

1. 开关. 即到达 S_i 的边界 C_i 后跳向另一流形 S_j 中的某个开集 B_i.

2. 跳跃. 即轨道到达 S_i 的某个区域 A_i 时跳跃到另一流形 S_j 的某开集.

上述开关或跳跃又可分为自主的(必然发生)或受控的(按要求发生)两种机制. 除传统的状态连续演化伴随着某些品质损耗外,开关和跳跃也都带有一定损耗. 其最优控制问题则要求出连续和切换的控制规律,使得总损耗最小.

对这类形式化的一般提法,当然还附加了一些条件,他们证明了最优轨道和最优控制的存在性. 其求解条件是最优值函数 $V(x,u)$应满足一个广义的

拟变分不等式，从而可运用若干研究过的求解方法. 目前还只能对某些最简单的情形找出有效的算法.

格子气自动机(lattice gas automaton)　一类具空间分布状态的自动机. 它是元胞自动机用于流体力学、统计物理等方面计算和模拟问题的一种具体化形式. 该计算模型将流体所处的时间和空间完全离散化，按物理定律规定离散流体粒子之间相互作用及迁移的规则，就构成一个格子气自动机. 格子气自动机的基本特征是其高度并行性和纯粹的布尔运算. 它为困难的非线性问题求解提供了一种方便的算法，从本质上保证了数值稳定性、可方便地处理边界问题，并将收敛速度提高了几个数量级.

格子气自动机最初于 1973 年由三位法国流体力学家提出，称为 HPP 模型. 1985 年对元胞自动机理论和应用的深入研究导致 FHP 模型的出现. 这些模型已被成功地用于研究流体力学的纳维-斯托克斯方程、多相流问题、电磁流体、相交、晶体增长等一系列问题. 近年它还被用于描述公路交通车流密度和拥塞的关系等问题.

符号动力学(symbolic dynamics)　一种非线性动态系统理论. 它研究由若干符号按一定演化规则形成的动态系统. 研究的主要问题是各种系统在相空间中的行为是否可以用这种粗粒度的量化来描述，具体形式则表现为按动力学规则出现的符号串的性质. 但目前还只是对若干简单的模型进行了详细研究.

考察由有限区间(不失一般设为[0,1])到其自身的映射，如图 1 的单峰映射或图 2 的锯齿映射. 在图 1 情形，用符号 L,C,R 分别表示相空间的左、中(峰值点)、右三个区域. 由任一初值 x_0 出发反复迭代，可以得到一串由 $x_1=f(x_0)$，$x_2=f(x_1)$，…组成的轨道以及与之对应的由 L,C,R 组成的符号串(参见“符号动态系统”). 众所周知，由这种单峰映射(有时亦称逻辑斯谛映射)可以产生稳定极限点、周期解，以及当 $f(x)$的参数变化时引起的倍周期分岔而形成混沌的极为复杂多样的行为模式. 这些模式均可由其“简化”的符号序列得到表现. 对于图 2 的映射，这点更加明显. 当人们用二进制表示[0,1]中的任意实数 x_i 时，锯齿映射只不过把 x_i 的二进制真小数的小数点向右移一位，并把整数项(可能变为 1)归零. 这样当 x_0 为有限位小数时，序列将在有限步达到稳定平衡点 $x_N=0$；当 x_0 为其他有理数时，其二进制表示为循环小数，从而对应极限集为稳定周期解；而当 x_0 为无理数时，对应为无限不循环的非周期轨道(类似于混沌).

符号动力学理论包括为所有符号序列的大小顺序建立排序规则，特别是由转折点(如图 1 中的 C)出发的符号序列(称为揉序列)有特别重要的意义；

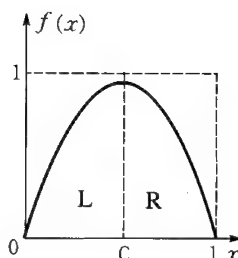


图 1　单峰映射

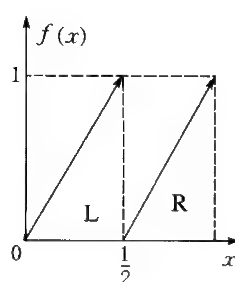


图 2　锯齿映射

研究有可能成为实际动力学轨道的符号序列应满足的条件，即允字条件；对其中混沌轨道的刻画，诸如由符号序列计算拓扑熵和比较复杂性的方法、确定出现混沌所对应的参数空间中的边界；计算某些类的符号字所对应的特征参数；产生允许字或对它们进行变换、合成的方法；在对称动力学系统中，由于参数改变而引起的对称破缺和恢复；诸如当参数变化时总共可能出现的周期行为之类的整体性问题等.

常被研究的映射包括一维情形的单峰映射，还有二次方映射、人字映射等多种形式；多转折点映射还有裂峰映射、立方型映射、正弦平方映射、洛伦兹型映射、圆映射等；二维相空间的映射研究有许多本质困难，现仅对特尔映射、埃隆映射、罗西映射等几种特例开展了研究.

符号动力学方法对研究动力学系统的重要性早在 1921 年就由莫尔斯(Morse,M.)指出，“符号动力学”的标题在他于 1938 年发表的论文中首次使用. 此后，鲍文(Bowen,R.)等人所做的重要进展于 1979 年在苏联成书出版. 米诺尔(Milnor,J.)和瑟斯敦(Thurston,E.)等人于 1977 年开始引入揉序列的概念，并阐明了其在研究符号动力学中的重要性.

有限自动机系统(finite automaton system)　最基本的一种离散型系统. 它可由五元组定义为 $S=(A,B,C,\phi,\lambda)$，其中 A 为输入符号有限集合(输入字母表)；B 为输出符号有限集合(输出字母表)；C 为有限状态集合；$\phi:C\times A\to C$ 为状态迁移函数；$\lambda:C\to B$ 为输出函数. 用 A^* 或 B^* 分别表示 A 或 B 中符号构成的符号串集合，则映射 ϕ 可扩张为 $\phi':C\times A^*\to C$，表示将 A^* 中输入串逐步作用于 C 中某状态而得到的最后状态. 有限自动机研究的基本问题是它如何接纳输入符号串 $a^*\in A^*$，并把初态 c_0 逐步转换为串 $c_0^*\in C^*$ 和输出串 $b^*\in B^*$. 由于通常把所有符号串集合的某一子集称为(形式)语言，已经证明，有限自动机系统所能接纳或识别的输入符号串和生成的输出符号串，只能是被称为正规语言的较小的集合. 上述 ϕ 和 λ 的定义亦可改为由 ϕ 和 μ

定义. 其中 $\phi: C\times A\to C$ 的意义同前，而 $\mu: C\times A\to B$ 直接给出下一个输出. 前者亦称为摩尔型自动机，后者则称为米雷型自动机.

摩尔型自动机(Moore automaton)　见“有限自动机系统”.

米雷型自动机(Mealy automaton)　见“有限自动机系统”.

L 系统(L-system)　一种用迭代方式模仿植物生长等演化过程的代数系统. 它由荷兰数学家林登迈尔(Lindenmayer, A.)于 1968 年提出. L 系统特点为：

1. 并行的重写过程.

2. 其语法可视为一种动态过程的描述，因此它开创了对其产生的词的序列(而不是集合)的研究.

在 L 系统语法中不存在终止字母. 例如，其中最基本的一类为 DOL 系统，它定义为三元组 $G=(\Sigma, h, \omega)$，Σ 为字母表，h 为定义于字母串集 Σ^* 上的自同态，ω 为公理. 由 G 产生的词序列 $E(G)$ 为：$h^0(\omega)=\omega, h^1(\omega), h^2(\omega), \cdots$，$G$ 的语言则定义为

$$L(G)=\{h^i(\omega)\mid i\geqslant 0\}.$$

L 系统亦可分为“上下文有关”、“上下文无关”等. 例如，字母表 $\{a,b,c,d\}$ 重写规则为：

1. $a\to cb$.

2. $b\to a$.

3. $c\to da$.

4. $d\to c$.

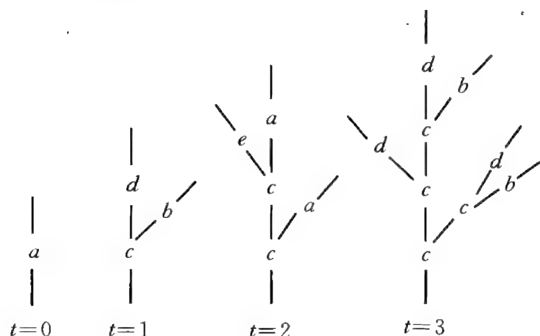
那么，由种子元素 a 开始的生长过程为

$$a,\quad cd,\quad daa,\quad ccbcb,\quad dadaadaa.$$

还可引入分岔，用(・)表示向左分岔，[・]表示向右分岔. 由重写规则 $a\to c[b]d, b\to a, c\to c, d\to c(e)a, e\to d$ 可得生长过程

$$a,\quad c[b]d,\quad c[a]c(e)a,$$
$$c[c[b]d]c(d)c[b]d, \cdots.$$

对应生长图形为：



不确定性系统

模糊系统(fuzzy system)　用模糊集概念描述不确定性的一种系统模型. 当把基于集合论的一般系统定义中系统各要素的论域上的集合均改为模糊集，就得到模糊系统. 它实质上是把系统中输入、输出及状态等各种变量加以模糊化的结果，特别适于描述含有模糊不确定性的社会经济、管理决策、人机交互、智能思维等人类活动系统，及人们尚不够了解其运行规律的自然界或人造的工程系统. 形式上，模糊系统 S 可定义为五元组：

$$S=\{X, U, Y, \delta, \beta\},$$

其中 X, U, Y 仍分别为状态空间、输入空间和输出空间，但其中的变量均可以是模糊子集. 状态转移函数 β: $X\times U\to X$ 和输出函数 δ: $X\to Y$ 亦均为模糊函数. 按离散时间演化的模糊系统的状态方程和输出方程，可分别写成 $x_{t+1}=\delta(x_t, u_t)$，$y_t=\beta(x_t)$，只是这里 x, u, y 分别是 X, U, Y 上的模糊子集，δ 和 β 则分别为 $X\times X\times U$ 和 $X\times Y$ 上的模糊函数，在某些情况它甚至可以是更一般的模糊关系. 和传统系统一样，对模糊系统的研究课题也包括稳定性和渐近行为、能控性、能观性、实现问题、系统辨识、预测和控制等方面.

模糊集(fuzzy set)　一种边界不分明的集合. 论域中的元素以不同程度隶属于模糊集并作为它的成员. 可将论域 $X=\{x\}$ 上的模糊集 A 定义为各种可能的 $\{x, \mu_A(x)\}$，其中 $\mu_A(x)$ 在 $[0,1]$ 取值，表示元素 x 对模糊集 A 的隶属程度，称为隶属函数. 当 $\mu_A(x)$ 只取值 0 或 1 时，就是常规集合的特征函数，从而模糊集是常规集合的一种推广. 日常生活中，新、老，大、小，高、矮等大量概念都是模糊的，它们依使用场合或人的主观感觉不同而变. 在工程技术系统中，误差大小是否可容忍，响应快慢或品质好坏是否令人满意，图形或文字判别时若干对象的差别，都是模糊的，自然语言和专家知识等更是典型的模糊对象. 因此，美国数学家扎德(Zadeh, L. A.)于 1965 年提出的模糊集理论具有很大的现实意义，并得到了普遍重视和迅速发展. 模糊集的各种运算可由普通集合运算进行自然的推广而定义如下：

1. 等价. 称模糊集 A 与 B 等价，即 $A\equiv B$，是指对任何 $x\in X$，均有 $\mu_A(x)=\mu_B(x)$.

2. 包含. 称模糊集 A 包含于 B 中，或称 A 是 B 的模糊子集，即 $A\subset B$，是指对任何 $x\in X$，均有

$$\mu_A(x)\leqslant\mu_B(x).$$

3. 补集. 模糊集 A 的补集记为 $\bar{A}$，定义为对任何 $x\in X$，有

$$\mu_{\bar{A}}(x)=1-\mu_A(x).$$

4. 并集. 模糊集 A 和 B 的并记为 $A\cup B$，定义为对任何 $x\in X$，有

$$\mu_{A\cup B}(x)=\max\{\mu_A(x), \mu_B(x)\}.$$

5. 交集. 模糊集 A 和 B 的交记为 $A\cap B$,定义为对任何 $x\in X$,有

$$\mu_{A\cap B}(x)=\min\{\mu_A(x),\mu_B(x)\}.$$

基于模糊集可以建立模糊系统的理论(参见"模糊系统"),其中最基本的概念之一是映射 $f:X\to Y$ 概念的模糊化扩展. 若 $y=f(x)$ 为单值函数,则自然有 $\mu_Y(f(x))=\mu_X(x)$. 若为多值函数,则定义

$$\mu_Y(y)=\max_{f(x)=y}\mu_X(x).$$

这就是扎德于 1975 年提出的扩展原理.

模糊集的非模糊化,亦称解模糊,也是应用中的一个关键问题. 模糊集 A 中隶属度不低于 λ 的元素构成的集合 $A_\lambda=\{x\mid\mu_A(x)\geqslant\lambda\}$,称为 A 的 λ 水平截集. 模糊集理论的分解定理指出,集 A 可以表示为

$$A=\bigcup_{\lambda\in[0,1]}\lambda A_\lambda,$$

其中 λA_λ 称为 λ 与 A_λ 之积,其定义为

$$\mu_{\lambda A_\lambda}(x)=\begin{cases}\lambda, & x\in A_\lambda,\\ 0, & x\notin A_\lambda.\end{cases}$$

应用分解定理可以把传统集合论中的许多基本关系推广到模糊集合论中去. 为解决某些特殊问题,还定义了若干特殊的模糊集合运算,如有界和、有界差、有界积、幂乘等.

模糊系统理论(fuzzy system theory)　一种基于模糊集概念而建立的系统理论. 它着眼于描述和处理不确定对象的模糊系统理论. 目前的成果主要是以模糊集合取代普通集合,把建立在精确数学基础上的系统概念、原理和方法推广到模糊系统上去. 另一条途径是模仿人脑思维的特点,建立新的问题求解的概念、原理和方法,并用于人类认知决策和智能活动的各个方面.

模糊性(fuzziness)　事物类属的不分明性. 在美国数学家扎德(Zadeh,L. A.)提出模糊理论之前,系统科学的研究对象具有或不具有某种属性,明确肯定,毫不含糊,从而在数学上,元素 x 属于集合 A 或不属于集合 A,非此即彼. 但日常生活中,新与旧、老与少、高与矮、大与小等大量概念所反映的对象属性没有明确的界限,事物的分类常依场合或人的主观感觉不同而变化. 在复杂工程问题中,误差大小、响应快慢、品质高低、决策优劣等也都难以给出精确的判别,通常只能加以定性的区分和处理. 这种现象反映在数学上,必须承认不同元素可以按不同程度属于某集合或不属于某集合. 总之,模糊性的数学描述是人类对事物认识和描述的一种飞跃. 是人类知识行为的一大飞跃.

模糊关系(fuzzy relation)　传统关系概念的推广. 论域 X 与论域 Y 间的模糊关系 R 是指直积 $X\times Y$ 上的模糊集合 R,由隶属函数 μ_R 描述,函数值 $\mu_R(x,y)$ 代表元素 x 和 y 具有关系 R 的程度. 若 $X=\{x_1,x_2,\cdots,x_m\}$ 和 $Y=\{y_1,y_2,\cdots,y_n\}$ 均为有限集,则其间模糊关系 R 可用 $\mu_R(x_i,y_j)$ 构成的 $m\times n$ 矩阵来描述,称为模糊关系矩阵. 同样概念可推广到 N 个论域 $X_1,X_2,\cdots,X_N$ 的元素之间的关系,它是 $X_1\times X_2\times\cdots\times X_N$ 中的模糊子集 R,用隶属函数 $\mu_R(x_1,x_2,\cdots,x_N)$ 来描述. 按模糊集间运算的定义,同样可以定义关系 R 的补集、关系 R_1 和 R_2 的并 $R_1\cup R_2$、交 $R_1\cap R_2$ 等. 除了常用的恒等关系 I、零关系 0、全称关系 E 和逆关系 R^{-1} 等概念均已简单地推广到模糊关系的情形,通常还要用到以下概念:

1. $X\times Y$ 上的模糊关系 R 与 $Y\times Z$ 上模糊关系 S 的合成 $R\circ S$:对任何 $(x,z)\in X\times Z$,均有

$$\mu_{R\circ S}(x,z)=\bigvee_{y\in Y}[\mu_R(x,y)\wedge\mu_S(y,z)],$$

其中 $\vee$ 和 $\wedge$ 分别表示"取极大"和"取极小"运算.

2. $X\times X$ 上的模糊相似关系 R. 对任何 $x,y\in X$ 均有 $\mu_R(x,x)=1$(自反性)及 $\mu_R(x,y)=\mu_R(y,x)$(对称性).

3. $X\times X$ 上的模糊等价关系 R. 除上述自反性和对称性外还满足传递性,亦即对任何 $x,y,z\in X$ 均有

$$\mu_R(x,z)\geqslant\bigvee_{y\in X}[\mu_R(x,y)\wedge\mu_R(y,z)].$$

在应用上经常会遇到模糊关系方程,亦即已知模糊关系 Q 和 S,求 R 使满足 $Q\circ R=S$ 的问题.

模糊逻辑(fuzzy logic)　一种非布尔逻辑. 用以研究模糊命题演算和模糊推理. 传统的布尔逻辑用 1,0 分别表示真、假,而模糊逻辑以 $[0,1]$ 中任意实数代表真或假的程度. 在相应的真值演算中,与($\wedge$,合取)、或($\vee$,析取)、非($\sim$)分别定义为

$$a\wedge b\triangleq\min(a,b),\ a\vee b\triangleq\max(a,b),\ \sim a\triangleq 1-a.$$

布尔代数的大多数公理均适用于模糊逻辑. 例外的只有补余律,即 $a\vee(\sim a)=1$ 和 $a\wedge(\sim a)=0$ 在模糊逻辑中不成立,这反映了模糊事物的亦此亦彼性. 在模糊谓词演算中,模糊命题的真值规定为它的隶属度:$T(P[u_0])\triangleq\mu_P(u_0)$,这里 $P[u_0]$ 表示语句"u_0 是 P". 这样,模糊谓词演算就可用前面定义的真值演算来确定. 类似地,还可定义模糊公式,即由上述运算构成的表达式,以及它们的合取、析取范式和最小化问题. 值得指出的是,最小化的结果与布尔代数不同. 因为补余律不再成立,从而 $x_i\wedge(\sim x_i)$ 在析取时不能删去,$x_i\vee(\sim x_i)$ 在合取时不能删去. 模糊推理亦称为似然推理,其中蕴涵关系 $a\to b$ 的定义有多种. 例如,在某些控制问题中就可采用 $a\to b\triangleq a\wedge b$(参见"模糊控制"). 模糊逻辑在预测、控制和识别上都有广泛应用,同时也推动了以此为基础的新型计算机软件和硬件的发展.

截集(cut set)　使用模糊描述非模糊化的一个基本概念. 在实际应用中,用模糊集合描述对象进行

推理之后，还要在最后阶段将它非模糊化，以给出确定的答案. 为此，需引入如下新的概念：模糊集合 A 的 λ 截集，记为 A_λ，是由 A 中一切隶属度不低于 λ 的元素构成的普通集合，即 $A_\lambda=\{x\mid\mu_A(x)\geqslant\lambda\}$. λ 称为置信水平，或称资格程度. 类似地，可以定义模糊关系 R 的 λ 截关系 R_λ. 基于截集概念提出的模糊截割理论，最后常归结为在一个适当的阈值上进行截割，并做出非模糊的判决.

不相容原理(inconsistency principle)　亦称互克原理. 关于模糊性和复杂性相互关系的基本命题. 这个原理断言：随着系统复杂性的增加，人们做出关于系统行为精确而有意义的陈述的能力将降低. 越过一定阈值，精确性和适合性将成为相互排斥的概念. 不相容原理表明对事物的精确描述和有意义的描述并不总是一致的. 精确的描述是手段而非目的，科学的描述必须能够反映对象的真实特性. 不相容性原理告诫人们，面对复杂的、特别是人的因素起重要作用的系统问题时，应放弃单纯追求高度精确化定量化，而应转向使用模糊的近似的、但能反映对象真实特性的描述.

模糊信息(fuzzy information)　反映事物模糊性或基于模糊集合刻画的信息. 仙农(Shannon，C. E.)提出的仅仅是语法信息，并不考虑其所传递的内容. 而人们日常传递、交流的大量都是模糊信息，对其真实性不能简单做出是或否的明确判断. 模糊信息的定量刻画涉及消息的模糊语义，还是一个有待研究的问题.

模糊控制(fuzzy control)　基于对模糊信息的描述和处理而提出的控制理论和技术. 传统控制的核心是建立精确的数学模型和算法，但对于那些过分复杂、机理不明、难以建立精确数学模型的系统，基于人的经验和模糊推理的模糊控制方法却可以取得令人满意的结果. 模糊控制的核心是：绕开建立精确数学模型而仿效人脑利用模糊信息进行模糊推理，依据人的经验制定模糊控制指令，从而简捷灵活地实现令人满意的控制. 建立模糊控制的算法模型大致有以下步骤：

1. 给出输入量和输出量的模糊量化与标定.
2. 建立模糊控制逻辑表.
3. 构造控制推理关系.
4. 设计输入、输出控制规则表，其基本形式为"若…，则…"的模糊条件语句，诸如，"若炉温偏高，则多吹入冷气"等.

模糊控制最初由英国人曼达尼(Mamdani，E. H.)等于 1974 年用于工业过程，其后又广泛应用于家电、工业、交通等领域，是一种前景广阔的新技术.

模糊聚类分析(fuzzy clustering analysis)　依据论域上的模糊等价(或相似)关系对论域中的对象进行分类的方法，它是传统聚类分析(参见《数学辞海》第四卷"聚类分析")把对象具体划分为若干等价类方法的推广. 模糊聚类分析的具体步骤为：

1. 把待分类的对象全体作为论域 X，建立从 X 到 X 自身的模糊等价关系矩阵 $R=(r_{ij})$，由于以 R 为分类标准得到的模糊关系往往不是等价的(不满足同时具有自返性、对称性和传递性要求)，而只是一种模糊相似关系(只具有自返性和对称性)，需再用平均法求出相似关系的传递闭包，求得的模糊等价关系即可以作为分类标准.

2. 取模糊等价关系 R 的 λ 截关系 R_λ，由于 R_λ 是一个以 0 和 1 为元素的普通等价关系矩阵，可由之按照传统方法进行分类. 令

$$R_\lambda=(r_{ij}),$$
$$r_{ij}=1\quad(r_{ij}\geqslant\lambda),$$
$$r_{ij}=0\quad(r_{ij}<\lambda).$$

把 R_λ 中元素相同的各行归为一类，得到一个关于论域中全体元素的分类. 不同 λ 值对应不同的分类结果. 根据问题给出的具体条件，确定一个关于聚类水平 λ 的序列：

$$1\geqslant\lambda_1>\lambda_2>\cdots>\lambda_k\geqslant 0,$$

按照 $R_{\lambda_1},R_{\lambda_2},\cdots,R_{\lambda_k}$顺序进行分类，得到一个由细变粗、逐步归并的聚类图. 这是一种动态的软分类，能够提供关于对象在不同聚类水平下不同类属关系的全面信息，适用于缺乏明确分类标准的复杂问题.

模糊模式识别(fuzzy pattern recognition)　基于模糊性概念的模式识别. 依据对象的模糊信息，按照模糊数学原理进行的模式识别. 常见的两种基本方法，即基于最大隶属度原则的直接方法和基于择近原则的间接方法. 直接方法适于解决的问题特点为：待识别的对象是明晰的，模式类型有模糊性，用模糊集合描述模式类型，识别任务是判明给定的对象应优先归属于哪个类别，或哪个对象优先属于给定的模式. 针对不同问题，直接法又可分两种情形：

最大隶属度原则 I. 给定论域上的模糊模式，用模糊集合 A 表示，论域中有 n 个待识别对象 $x_1,x_2,\cdots,x_n$. 识别任务是确定哪个对象优先归属于 A. 答案为：若 $\mu_A(x_i)=\max\mu_A(x_j)$，则优先将 x_i 归属于 A.

最大隶属度原则 II. 给定论域 X 上的 n 个模式，分别用模糊集合 $A_1,A_2,\cdots,A_n$ 表示，$x_0\in X$ 为待识别对象. 识别任务是确定 x_0 优先归属于哪于模式. 答案为：若 $\mu_{A_i}(x_0)=\max\mu_{A_j}(x_0)$，则 x_0 优先归属于 A_i.

间接方法要处理的问题是：模式类型和被识别的对象都是模糊的，都需用模糊集合表示，模式识别在数学上归结为衡量两个模糊集合的接近程度，按

照接近程度决定对象归属于哪个模式.常用的数学工具是贴近度和择近原则.模糊集合 A 和 B 的贴近度记为 $\xi(A,B)$,有不同定义,需根据实际问题而选定.择近原则也有两种表述的形式:

择近原则 I.给定论域上的一个模糊集合 A,代表一个模糊模式,论域中的 m 个待识别对象分别用模糊集合 $B_1,B_2,\cdots,B_m$ 表示.识别任务是确定哪个对象优先属于 A.答案为:若

$$\xi(A,B_i) = \max \xi(A,B_j),$$

则把 B_i 优先划归于 A.

择近原则 II.给定论域上的 m 个模糊集合 $A_1,A_2,\cdots,A_m$,代表 m 个模糊模式,被识别对象为模糊集合 B.识别任务是确定 B 应优先归属于哪个模式.答案为:若

$$\xi(A_i,B) = \max \xi(A_j,B),$$

则把 B 优先归属于 A_i.

模糊综合评判(fuzzy comprehensive judgement) 模糊数学在决策分析中的一种应用.在模糊环境中,运用模糊集合概念权衡各种因素,给对象优劣以综合评价的方法.决策目标为对各待选方案的优劣做出评判.评判活动由待评估因素(性能指标)和评语两方面构成.由于一般涉及多种因素,必须进行综合评判.因素和评语一般都有模糊性,需用模糊集合和模糊关系(矩阵)刻画待评估因素和评语,并用相应的运算描述评判活动,并做出带有模糊性的结论,这就是模糊综合评判.

模糊决策(fuzzy decision) 着眼于描述和处理模糊性而提出的决策理论和方法.在只能依靠模糊信息进行决策的情况下,人们通常的想法是:如果系统大体为某一状态时,大体采用某种行动方案就比较好.利用模糊数学进行这种决策,就是模糊决策.在许多情形下,只要按各种可能方案进行排序,或按照某种方法从中选择较优者,就算达到了决策目的.模糊综合评判就是一种模糊决策方法.

模糊优化(fuzzy optimization) 目标或约束为模糊变量的优化问题.处理模糊优化的基本原理是:将模糊目标、模糊约束均作为解集合上模糊子集处理,用隶属函数表示这两个模糊集合,求取模糊目标和模糊约束的交集,则交集隶属函数的最大化,就是该模糊优化问题的最优解.

粗糙集(rough set) 一种具不确定性的集合概念.它基于对经典集合论的扩展,把知识嵌入到集合内,基于分析对象的不可分辨关系,揭示知识在论域内的颗粒状结构.粗糙集理论是波兰学者波拉克(Pawlak,Z.)于1982年提出的.粗糙集理论能有效地处理不精确知识表达、知识获取、知识推理,广泛应用于机器学习、模式识别、知识发现、数据挖掘等领域,是人工智能领域研究的热点之一.在粗糙集理论中,知识被理解为一种分辨能力.系统 S 的知识表示可以用四元组表示:$S=(U,A,\{V_a\},\{a\})$,式中 U 为论域,A 为属性集合,V_a 为属性的值域,a 为 $U\to V_a$ 的单一映射.在分辨过程中,把相差不大的个体归为同一类,它们的关系称为不可分辨关系,简记为 IND.即对于知识表示系统 S,有 $B\subseteq A$,把二元关系

$$\mathrm{IND}(B)=\{(x,y)\in U\times U;\ \forall a\in B,\ a(x)=a(y)\}$$

称为 S 的不可分辨关系,式中 x,y 为 U 中的元素.不可分辨关系又称为等效关系,简记为 ER.

论域中相互不可分辨的个体组成的集合,称为基本集,或称基类.显然,不可分辨关系就是等价关系,它满足自反性、对称性和传递性.通过这种不可分辨关系得到的所有基本集,实际上形成了在论域上依据知识对问题的划分.也就是说,基本集组成了论域知识的颗粒.当人们的知识越多,则对问题认识越细,基本集的颗粒就越小;反之则对问题的认识就比较粗略,基本集的颗粒就较大.所以,在粗糙集理论中,不可分辨关系揭示了知识的颗粒状态结构.总之,知识可以认为是一簇等效关系,它将论域分割成一系列等效类.

设 I 为 U 上的一簇等效关系,即 I 是关于论域 U 的知识,则二元对 $K=(U,I)$ 称为近似空间.设 x 是 U 中的一个对象,X 为 U 的一个子集,$I(x)$ 表示所有与 x 不可分辨的对象组成的集合,或者说是由 x 决定的等效类,即 $I(x)$ 中的每个对象都与 x 具有相同特征属性.由于系统具有不定性,在近似空间上又作如下定义:

$$\underline{I}(x) = \{x \in U, I(x) \subseteq X\}$$

定义为集合 X 关于 I 的下近似;

$$\overline{I}(x) = \{x \in U, I(x) \cap X \neq \varnothing\}$$

定义为集合 X 关于 I 的上近似.$\underline{I}(x)$ 实际上是由那些根据现有知识判断肯定属于 X 的对象所组成的最大集合,又可以称为 X 在 I 下的正区域,记为 $\mathrm{POS}(X)$;而 $\overline{I}(x)$ 实际上是由那些根据现有知识判断肯定不属于 X 的对象所组成的集合,它又可以称为 X 在 I 下的负区域,记为 $\mathrm{NEG}(X)$.上近似与下近似之差称为边界区域,定义为

$$\mathrm{BND}(X) = \overline{I}(X) - \underline{I}(x).$$

如果 $\mathrm{BND}(X)$ 是空集,则称 X 关于 I 是清晰的;反之,如果 $\mathrm{BND}(X)$ 不是空集,则称集合 X 为粗糙集.

灰色系统理论(grey system theory) 一种处理信息不完全问题的系统理论.它以信息不完全、不确定系统的行为表现、行为内涵、行为关系、行为环境的层次性、动态性、信息性、量化性等为研究对象.由中国学者邓聚龙于1982年提出.灰色系统理论的主要内容有:

1. 灰观念.包括认识无穷尽公理、灰性不灭原

理、层次无限可分原理、解的非惟一性原理、差异信息原理、新息优先原理、信息传递模式、认用模式和结构模式.

2. 灰生成. 包括数据的数值变换、数据的层次转换和互补律的数字处理.

3. 灰关联. 建立整体比较机制，收取距离空间的量化特性与手段，融合点集拓扑领域的整体比较内涵，升华为灰关联空间. 在灰关联空间中，可以确定因子的序化关系，划分系统主部分，评估因子对主行为贡献的大小.

4. 灰建模. 在采集序列的基础上，建立近似微分方程模型，简称灰模型，记为 GM. 微分方程以连续可微函数为背景. 作为序列，"连续"与"可微"条件是不具备的. 为此，灰色系统理论从序列的角度出发，剖析一般微分方程的构成条件，然后对那些近似满足条件的序列建立近似的(灰的)微分方程模型. 灰色建模需要的数据少，只要有四个(或四个以上)数据就可以建立灰色模型，并且允许序列中第一个数据为灰数.

5. 灰预测. 以 GM(1,1)模型为工具，对事物的时间分布、数值分布进行预测，称为灰预测. 它包括：数列预测、突变预测、季节灾变预测、拓扑预测和系统预测.

6. 灰决策. 主要分为灰局势决策、灰层次决策、灰规划等. 其中灰局势决策是测度空间的决策. 测度空间是目标极性一致化的空间，测度是目标样本的抽象. 在测度空间中，"测度大"可以代表"效益大"、"损耗小"、"样本大小适中". 总之，"测度大"说明接近目标的程度大. 测度的转换内涵，体现多目标到单目标的转换.

7. 灰控制. 灰色控制，是基于灰色模型 GM(1,1)的控制. 建模只需单序列，即行为序列. 控制过程中的每个采样时刻，根据一个新的采样值，按"新陈代谢"原则建立一个新的 GM(1,1)模型，从而得到一组新的模型参数和预测值，以进行新的实时控制，这实际上是用行为变量和模型参数变化来预测控制量度，以适应各种干扰、环境及本身的变化，并保持控制的精度.

从灰色系统理论所涉及的问题和解决问题的方法看，它有以下特点：

1. 可用较少数据，例如可用三数据加一灰数建模.

2. 运用较少数据分析，例如三数的灰关联分析.

3. 近似(灰)微分方程模型的建立.

4. 高阶微分方程求解问题. 对各阶微分方程，灰色系统理论找到了一种介于解析解与数值解之间的解.

5. 噪音与系统主行为的认同. 灰色过程是行为表现过程，不是因果过程，不必区分噪音表现与非噪音表现，可回避处理随机过程的麻烦.

6. 多维灰色规划. 既考虑规划的动态性、灰色性，又考虑规划的综合性.

灰色系统理论已形成了以灰关联空间为基础的分析体系，以灰模型 GM 为主体的模型体系，以灰色过程、生成空间为基础的内涵方法体系，以系统分析、建模、预测、决策、控制、评估为网的技术体系，并广泛应用于农业、经济、社会、生态、气象、医学、管理、政治、环境等领域，取得显著成效.

并行计算系统

并行计算系统(parallel computing system) 高速计算和数据处理系统的一种实现形式. 它由分布式的多台计算机和高速网络互联，并行执行同一计算任务以提高速度和效能. 与拥有多个处理器和紧密耦合内存的巨型计算机方案相比，并行分布式计算系统可以把具有不同操作系统平台的工作站连结在一起，以实现并行计算功能，故又称为集群工作站. 并行计算的本质要求不设单独的总控制工作站，这就要求研究负载在工作站网络间的自适应平衡分配. 集群工作站还可通过个体的学习、交互以提高其性能. 总之，集群工作站实现的关键在于将计算任务平衡分配到各工作站，并通过高速网络协调，在集群工作站内部实现自动规划，而对外部表现为一个整体. 其具体内容包括高速网络技术、并行算法研究、通信与同步技术、计算机体系结构等.

感知机(perceptron) 一种神经元网络模型. 由美国学者罗森布拉特(Rosenblatt, F.)于 1957 年提出. 它是一个单层网络，输入信息加权后被直接送到输出节点，经过阶梯函数作用后，给出输出信息. 原始的感知机算法只有一个输出节点，相当于单个神经元，其结构如图 1 所示.

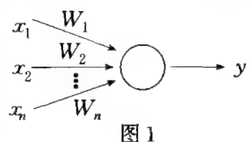


图 1

当用于两类模式的分类时，它相当于在高维样本空间，用一个超平面将两类样本分开，如图 2 所示，给出一组分别属于两类模式 A，B 的样本，当样本属于 A 类时，期望输出为 1，当样本属于 B 类时，输出为 -1，即实际的输出为

$$y=f\Big(\sum_{i=1}^{n}W_i x_i-\theta\Big),$$

$$y=\begin{cases}1 & (\text{属于 } A \text{ 类}),\\ -1 & (\text{属于 } B \text{ 类}).\end{cases}$$

$$y=f\Big(\sum_{i=1}^{n}W_i(t)x_i(t)-\theta\Big).$$

权系数调整的规则设计为

$$W_i(k+1)=W_i(k)+\eta(d(t)-y(t))x_i(t),$$

其中 $\eta\geqslant 0$ 是增益，$d(t)$ 是期望的输出，$y(t)$ 是实际的输出. 重复调整，直到权系数不发生变化为止. 多层感知机模型改变了原始单层模型的不足，使得感知机模型的应用更加广泛.

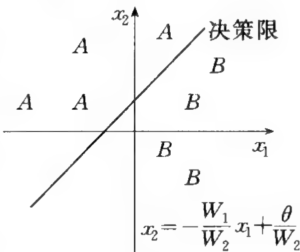


图 2

人工神经元网络(artificial neural network) 简称神经元网络. 模拟神经元计算机制的一种大规模计算和信息处理网络. 最早由美国心理学家麦卡洛克(McCulloch, W.)和数学家匹兹(Pitts, W.)于 1943 年提出，称为 MP 模型；1949 年，赫伯(Hebb, D.)提出了神经元之间连结强度变化的学习规则，称赫伯规则，从而开创了神经元网络研究的新局面. 神经元网络是由大量类似于神经元的处理单元相互连结而成的复杂非线性网络系统，它试图通过模拟大脑神经网络存入、处理信息的方式，部分地实现人脑信息处理的功能. 尽管每个神经元的结构和功能十分简单，由大量神经元互连组成的网络系统会表现出极为复杂的行为和极强的信息处理功能. 诸如：大规模并行处理，分布式存贮，极强的适应性、鲁棒性和学习能力等.

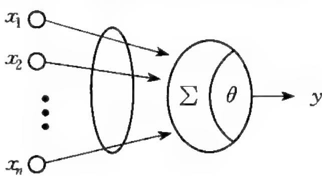


图 1

人工神经元网络的基本处理单元是神经元，其结构如图 1 所示. 其中，$x_1,x_2,\cdots,x_n$ 是神经元接收到的信息，y 是神经元的输出信息，其间关系为

$$y=f\left(\sum_{i=1}^{n}W_ix_i-\theta\right),$$

其中，$W_i(i=1,2,\cdots,n)$ 是互连权值，$f(x)$ 是非线性函数，它的形式有很多种，主要的几种形式见图 2.

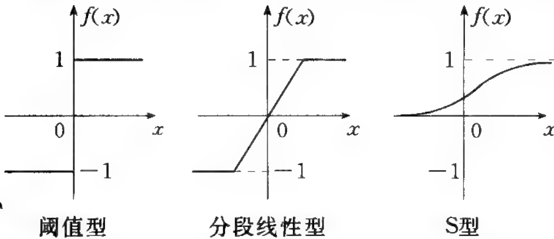


图2

人工神经元网络理论和应用已渗透到很多领域，并在智能控制、模式识别、非线性优化、计算机视觉、人工智能、信息处理、传感技术、机器人、故障检测与诊断以及生物医学等方面取得了很大进展. 研究神经元网络的模型主要有霍普菲尔德模型、反向传播模型、感知机、MP 模型、柯含农自组织模型等.

麦卡洛克-匹兹模型(McCulloch-Pitts model) 简称 MP 模型. 一种早期的神经元网络模型. 由美国心理学家麦卡洛克(McCulloch, W.)和数学家匹兹(Pitts, W.)于 1943 年共同提出. 设有 n 个神经元相互连结，每个神经元的状态 $S_i(i=1,2,\cdots,n)$ 取值 0 或 1，分别表示该神经元的抑制和兴奋，每个神经元的状态都受其他神经元的制约，

$$S_i=\sigma\left(\sum_j W_{ij}S_j-\theta_i\right)\ (i=1,2,\cdots,n),$$

$\sigma(x)$ 是阶梯函数，

$$\sigma(x)=\begin{cases}1 & (x\geqslant 0),\\ 0 & (x<0);\end{cases}$$

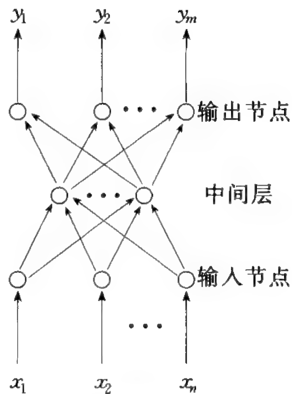
θ_i 是第 i 个神经元的阈值，W_{ij} 是神经元 i 与神经元 j 之间的连结强度，是可调的实数. 学习就是调整 W_{ij} 的过程. MP 模型的学习过程是按照赫伯(Hebb, D.)规则进行的，其原则为：若 i 和 j 两个神经元同时处于兴奋状态，则它们之间的连结应加强，否则削弱，即

$$\Delta W_{ij}=\alpha S_iS_j\ (\alpha>0).$$

这一规则与"条件反射"学说是一致的，并已得到细胞学研究的证实.

反向传播模型(back propagation model) 亦称 BP 算法. 一种多层前馈神经元网络的学习算法. 由鲁梅哈特(Rumelhart, D.)等人于 1985 年提出. 这个模型含有输入节点、输出节点及一层或多层隐节点，其结构如图所示. 输入信号先向前传播到隐节点，经过作用函数的运算，把信息送到输出节点，得到一个输出，然后让它与期望的输出比较，如果偏差满足规定的要求，那么学习过程就结束，否则网络就要从输出节点反向传播，逐层修改权系数. 节点的作用函数通常选用 S 型函数，如

$$f(x)=\frac{1}{1+e^{-x}}.$$



BP 模型

BP 算法首先从分析偏差的平方和入手，对第 p 个样本，要求误差

$$E_p=\frac{1}{2}\sum_j(\overline{y}_{pj}-y_{pj})^2\rightarrow\min,$$

其中 $\overline{y}_{pj}$ 为期望的输出，y_{pj} 为实际的输出. 信号正向传播过程中，

$$y_{pj}=f_j\left(\sum_i W_{ji}\cdot I_{pi}\right).$$

BP 算法利用梯度下降法对权值进行修正，即

$$\Delta_p W_{ji} \propto -\frac{\partial E_p}{\partial W_{ji}},$$

经一系列计算可得：

学习算法： $\Delta_p W_{ji}=\eta \cdot \delta_{pj} \cdot I_{pi}$；

对于输出节点：

$$\delta_{pj} = (\bar{y}_{pj} - y_{pj}f'_j) \sum_j W_{ji}I_{pi},$$

对非输出节点：

$$\delta_{pj} = f'_j\left(\sum_i W_{ji}I_{pi}\right) \cdot \sum_k \delta_{pk}W_{kj}.$$

霍普菲尔德模型(Hopfield model) 一种人工神经元网络模型，由霍普菲尔德(Hopfield，J. J.)于1982年提出．它包括离散和连续两种模型．离散霍普菲尔德模型是一种离散时间网络，它由 n 个神经元构成．S_i 取值为1或－1，神经元随机、异步地按下述规则改变状态：

$$S_i=\begin{cases}1 & \left(\sum_j W_{ij}S_j-\theta_i\geqslant 0\right), \\ -1 & \left(\sum_j W_{ij}S_j-\theta_i< 0\right).\end{cases}$$

对于 $W_{ij}=W_{ji}$ 的对称互联网络，引入能量函数(李亚普诺夫函数)

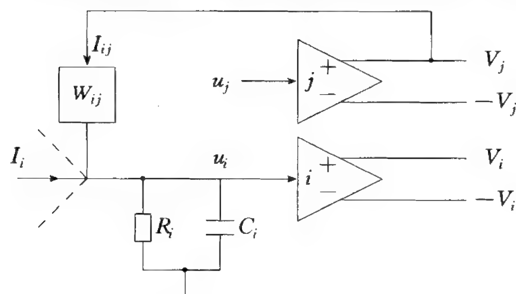
$$E = -\frac{1}{2}\sum_{j\neq i} W_{ij}S_iS_j + \sum_i \theta_i S_i,$$

则网络演变过程 E 是单调下降的，即

$$\Delta E = \Delta S_i \cdot \frac{\partial E}{\partial S_i} = \Delta S_i\left(-\sum_{j\neq i} W_{ij}S_j + \theta_i\right) \leqslant 0,$$

所以系统将趋于定态．

连续霍普菲尔德模型是连续时间系统，如图所示，其中电阻 R_i 和电容 C_i 并联以模拟神经元输出的时间常数；跨导 W_{ij} 模拟神经元突触之间的互联特性；运算放大器模拟神经元的非线性特性．u_i 是第 i 个神经元的输入，V_i 为其输出．设有 n 个神经元互联，则可用下述非线性微分方程描述：



霍普菲尔德神经元

$$\begin{cases}C_i \dfrac{\mathrm{d}u_i}{\mathrm{d}t}= \sum_{j=1}^{n} W_{ij}V_j-\dfrac{u_i}{R_i}+I_i, \\ V_i=g(u_i) \quad (i=1,2,\cdots,n),\end{cases}$$

其中

$$\frac{1}{R_i} = \theta_i + \sum_{j=1}^{n} W_{ij}, \quad V_i = g(u_i)$$

为神经元的非线性特性．对于连续的对称互联网络，它的能量函数

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}V_iV_j - \sum_{i=1}^{n} V_iI_i + \sum_{i=1}^{n}\frac{1}{R_i}\int_0^{V_i} g^{-1}(V)\mathrm{d}V.$$

同离散模型一样，网络的演变总是朝着能量 E 减小的方向运动，即

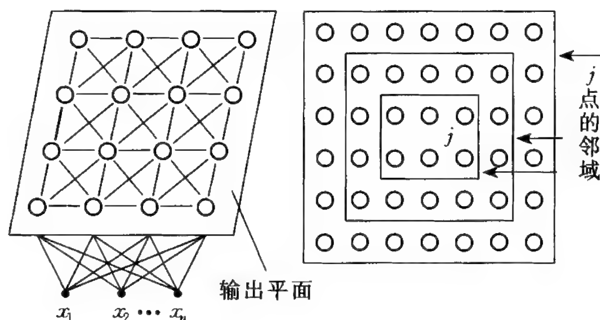
$$\frac{\mathrm{d}E}{\mathrm{d}t} \leqslant 0.$$

霍普菲尔德网络的权值可以根据具体问题而定，一般情况下

$$W_{ij}=\begin{cases}\sum_{s=1}^{n} x_i^s x_j^s & (i\neq j), \\ 0 & (i=j).\end{cases}$$

霍普菲尔德用这个模型对著名的“旅行推销员”、NP完全问题进行了讨论．

柯含农自组织模型(Kohonen self-organization model) 一种聚类分析的神经元网络模型．由芬兰学者柯含农(Kohonen，T.)于1972年提出．该模型可在一维或二维的处理单元阵列上给出各种输入模式的自组织分类图．其结构示意图如图所示．它的自组织特征映射功能由以下四部分组成：



1. 处理单元阵列，它由事件空间接受输入，并形成这些阵列的初始辨识函数．

2. 比较选择机制，它对辨别函数进行比较，并选出具最大输出函数值的单元．

3. 局部互联作用，它可同时激励被选出的处理单元和最近的邻接处理单元．

4. 自适应过程，它可修正被激励处理单元的参数，以增强其相应于特定输入的辨别函数输出值．

这种自组织模型具有抽取输入信号模式特征的能力，因而它可广泛用于模式识别、语音和图象特征提取、匹配和识别等领域．

联想记忆(associative memory) 神经元网络重要的功能之一．人们认识客观世界是通过事物之间的联系，由一事物联想到另一事物，并用联想记忆的方式加以固定．联想记忆是一个输入向量 $x_i \in \mathrm{R}^n$

到输出向量 $y_i \in \mathrm{R}^m$ 的映射，即

$$M: x_i \rightarrow y_i \ (i=1,2,\cdots,n).$$

联想记忆的过程分为两个阶段，即学习阶段和联想回忆阶段. 在学习阶段，各种被归纳好的样本存贮到网络中；在联想回忆阶段，当一个信号输入时，网络根据自身的算法，得到相应的输出. 联想记忆的方式分为自联想记忆和异联想记忆两种. 自联想记忆的学习过程中，事先存入 n 个样本 $x_i(i=1,2,\cdots,n)$，当有一信号 $x=x_i+\Delta$ (Δ 为偏差，可以是干扰、噪声等)输入时，网络就可以输出 $y=x_i$，达到联想自身的目的. 而在异联想记忆学习过程中，事先存入具有对应关系的两组样本 $x_i \rightarrow y_i(i=1,2,\cdots,n)$，当输入一个信号 $x=x_i+\Delta$ (Δ 的意义同上)时，网络可以输出 $y=y_i$. 显然，自联想记忆是异联想记忆的一个特例，后者的适用范围更广.

模拟退火(simulated annealing) 简称 SA. 寻找大规模优化问题全局近似解的一种随机搜索技术. 其原理来自于统计力学. 20 世纪 80 年代开始被用于求解组合优化问题. 采用这种技术应包含以下三个步骤：

1. 建立与物理概念之间的对应关系. 能量函数对应于优化的目标函数；物理系统的一个微观状态对应于参数空间的一个状态；最低能量的状态对应于最优解；温度对应于过程的控制参数.

2. 设定退火表，即设定一组下降的温度序列 $T_i(i=1,2,\cdots,n)$，以及在每个温度 T_i 上的停留的时间.

3. 给出生成和选择新的状态的规则.

模拟退火的基本原理：在温度 T 时，达到热平衡后一个微观状态 s 出现的概率由玻耳兹曼分布给出

$$q_T(s)=\frac{\mathrm{e}^{-\frac{E(s)}{kT}}}{\sum_{w\in S}\mathrm{e}^{-\frac{E(w)}{kT}}}, \qquad (1)$$

其中，k 为玻耳兹曼常数，$E(s)$ 为相应于微观状态 s 的能量，S 为全部可能的状态的集合. 因此，随着温度 T 的下降，相应于低能量的状态出现的概率将增大. 当一个新的状态 s_1 产生后，相应的 $q_T(s_1)$ 如果大于 $q_T(s)$ 时，状态 s_1 将自动被接受，反之状态 s_1 被保留的概率为

$$\frac{q_T(s_1)}{q_T(s)}=\mathrm{e}^{-\frac{E(s_1)-E(s)}{kT}}. \qquad (2)$$

为了达到最低能量的状态，不能简单地降低温度，应该使用退火过程，即升高系统的温度，然后逐渐降低. 并且要有充足的时间，使得在每一温度下能够达到热平衡.

演化计算(evolutionary computation) 一类模拟自然进化过程的算法. 它包括遗传算法、演化规划、演化策略. 这三种方法都是一类借鉴生物界自然选择和自然遗传机制的随机搜索算法，它们有两个共同的特点：群体搜索策略和群体中个体之间的信息交换. 这两个特点的优越性表现在：首先，演化算法在搜索过程中不容易陷入局部最优，即使在所定义的适应度函数是不连续的、非规则的或有噪声的情况下，它们也能以很大的概率找到全局最优解；其次，由于它们固有的并行性，演化算法非常适合于巨型并行机；再者，演化算法采用自然进化机制来表现复杂的现象，能够快速可靠地解决非常困难的问题；此外，它们容易介入到已有的模型中并且具有可扩展性.

遗传算法(genetic algorithm) 简称 GA. 一种全局随机寻优算法，由美国圣菲学派霍兰(Holland, J.)于 1975 年提出. 在 GA 中，用于表示各类复杂结构数据的基本形式是所谓的染色体(或者说用二进制 0,1 表示的数串). 一个典型的 GA 由五个基本部分组成：

1. 问题解的染色体表示.
2. 产生初始解群体的方法.
3. 评价函数.
4. 产生子代染色体的遗传算子.
5. 设定算法过程中的各种参数，如群体规模、遗传算子的使用频数等.

遗传算法具有并行的特征. 其中交叉算子是 GA 之所以有效的关键. 直观上，GA 是对自然进化过程的一种类比. 通过遗传算子，从父代染色体产生子代染色体，经过"环境"的选择，以取得具有更强适应性的新结构. GA 中"环境"的选择作用是通过评价函数来表征的.

数据处理组合算法(group method of data handling) 简称 GMDH. 一种数据处理建模方法. 由苏联学者伊瓦赫年柯(Ивахненко, Г. А.)提出，是一种对自然过程归纳自组织的建模途径. 其中，特别是生物种群的遗传过程以及罗森布拉特(Rosenblatt, F.)提出的人脑信息处理模型对 GMDH 的产生具有重要影响. GMDH 所处理的典型问题是根据观察数据计算出一个恰当的多元非线性函数模型

$$y=f(x_1,x_2,\cdots,x_n).$$

此方法的特点是建模者能够把已有的一些经验和知识，以一种直接的方式参与建模的过程. 其依据则在于构成 GMDH 基础的整体随机寻优算法具有一种"开放性"特征. 目前，GMDH 在工程、环境、经济等方面都获得了富有成效的应用. GMDH 可以看成是遗传算法的一种特定方式，有时也被称为自组织建模方法.

数据处理组合算法的基本过程可概括为：

1. 收集所研究问题的样本数据，其容量为 N.

2. 将此 N 个数据分成两个集合 N_A，N_B，其中 N_A 是培训集，N_B 是检验集.

3. 在输入和输出变量间建立一个“参考函数”.

4. 在标准的规则集中选择合适的目标函数，如偏差最小一致性原则等.

5. 以参考函数为基础将不同的局部方程分类.

6. 用测试集 A 和参数估计的方法对各局部方程的权重进行估计.

7. 用测验集 N_B 根据选定的目标函数来对这些局部方程进行量度.

8. 选择最好的测量方程作为最佳模型.

元胞自动机(cellular automata)　一种空间分布的复杂系统模型. 最早由美籍匈牙利数学家冯·诺伊曼(von Neumann, J.)于 1948 年研究具有自我复制能力的机器人时提出. 它以空间中规则分布的大量独立元胞及其与少量邻近元胞的局部相互作用确定的演化规则来决定其动态过程及其全局行为.

模型的空间分布形式可以是 n 维空间中的立方格点、三角或六角形结构等. 最简单的情况是一维直线上等距离分布的元胞，这时可以用一个标号 $i=\cdots,-2,-1,0,1,2,3,\cdots$ 来指明，其宽度即元胞数量只要足够大，即可呈现出其宏观行为模式. 其次，每个元胞的状态 x_i 可取有限 k 个值，最简单的是两个，通常用 0 和 1 或作图时用黑白两种颜色表示. 第三，演化规则取决于与它相关的邻近元胞(包括其自身)的数量，例如 N 个. 元胞 i 在下一时刻 $t+1$ 的状态将这 N 个元胞的 t 时刻的状态完全决定. 易证，这里不同演化规则的种类数为 k^{k^N}. 当 $k=2$，$N=4$ 时，演化规则多达 65536 种，若改为 $N=8$，则约有 10^{27} 种之多!

一维直线上 $k=2$ 的元胞自动机实际上是个布尔化的离散动力系统. 若沿空间轴在每个元胞的方格中用黑色表示 1，用白色表示 0，再由某组给定初值出发，沿与之垂直的时间轴上不断标出这种黑白颜色的变化，就可以得到一种平面方格上黑白相间的图案，即元胞自动机宏观行为的一种表示.

更细致的研究可以发现，元胞自动机也会呈现对初值敏感，或某些复杂运动模式经过若干不规则过程后又重复再现的基因现象等复杂行为. 某些元胞自动机可以生成酷似自然界蛤壳纹理或雪花的图案，从而可能有助于探索自然界自组织生长机制的奥秘.

元胞自动机模型可以用于计算和仿真. 例如，用规模为 $20\times20\times20$，$k=4$ 的三维元胞自动机，适当设计演化规则，可以显示化学反应中浓度随时空变化振荡的反应波. 元胞自动机的混沌行为可以用于设计一种新的随机数发生器. 在元胞自动机中加入概率机制，可能导致系统状态的不连续变化，从而用于研究非平衡相变.

生命游戏(life game)　一种模拟生命现象的模型. 由康维(Conway, J.)于 1970 年提出. 它实际上是一种二维的元胞自动机，其每个元胞状态取布尔值 0 或 1，而下一状态依据其八个邻近元胞(称为摩尔邻域)的状态由下述规则确定：

1. 若有两个邻近元胞取值 1，则其状态不变.

2. 若有三个邻近元胞取值 1，则其新状态为 1.

3. 任何其他情形，其新状态均为 0.

这些规则的社会生态含义可解释为：若元胞和环境过于拥挤或孤单，则它将灭亡(规则 3)；只有当拥挤度适中时，它才能存活或衍生(规则 1 和 2). 康维的问题是该系统是否会出现自我复制过程? 若空间为无穷大，这种过程能否无限持续下去?

人们不难从不同初条件得到各种演化行为，包括各元胞逐渐灭绝、元胞分布趋于稳定或趋于周期振荡，或者保持相同形状而不断“爬行”或“滚动”等. 这就对康维的问题给出了符合直觉的答案. 人们还进一步发现了更多新奇的行为，引起一股热潮，推动了 20 世纪 80 年代元胞自动机的研究.

系 统 工 程

系统工程(system engineering)　按系统观点解决组织管理问题的技术. 是一种从整体出发，运用系统思维进行合理的开发、设计、实现、反馈、修正，直到达到符合目的要求的一项综合性工程技术. 它根据总体的需要，综合应用从自然科学到社会科学的相关跨学科思想，以电子计算机等为工具，采用运筹学等方法分析与协调系统的结构、要素、信息、反馈等诸多方面之间的相互联系，力求达到全系统最优的规划、设计、管理和控制的目的. 系统工程属系统科学的工程技术层次，旨在按明确目标对系统进行改造并直接解决这种工程技术问题组织管理的理论和方法. 系统工程在 20 世纪 50 年代由军事运筹方面推动而开始形成. 1950 年开始出版第一批专著，论述其理论和方法论问题. 中国科学家钱学森等人在 20 世纪 60 年代国防科技研究中成功运用系统工程方法取得丰富成果，并于 1978 年起发表了一系列重要的总结性文章，推动了系统工程在许多重要领域的成功应用，诸如：

1. 工程系统，特别是大型工程项目的规划、设计、制造和运行.

2. 社会系统，包括国际关系、国家和社会管理的有关问题.

3. 经济系统.

4. 农业系统.

5. 企业系统.

6. 科学技术管理系统.

7. 军事系统.

8. 全球环境生态系统.

9. 人口系统.

10. 交通运输系统.

11. 能源系统.

12. 区域规划系统.

13. 信息系统等.

系统分析(system analysis) 按预定要求对系统进行定性或定量分析的技术. 对此有两种理解：广义系统分析是系统工程的同义语，狭义系统分析则是系统工程的一个组成部分，它与系统设计、系统实施一起构成系统工程工作步骤的三个阶段，也是霍尔三维结构中逻辑维的重要内容(参见“霍尔三维结构”). 系统分析应用多学科领域知识和建模仿真等技术，对系统各方面进行定性和定量分析，再综合考虑成本、效益等因素，建立计算和数学模型，并在计算机上进行仿真实验. 在此基础上进行综合比较，并按特定目标选择多种方案，从而为决策提供科学依据. 系统分析的基本要素包含：

1. 系统目标、功能、技术条件和约束，以及可行性分析.

2. 基本途径和备选方案，对之进行分析、比较和优选.

3. 建立物理模型、数学模型、计算机仿真模型及它们的混合模型，进行实验和分析比较；

4. 费用成本分析.

5. 从社会、环境、经济、可靠性等多方面进行效益的综合评价，这种评价过程通常还要经过多个步骤以及它们的反馈修正，以期达到最佳的效果.

系统建模(system modelling) 对给定系统按待研究问题建立模型的原理和技术. 它针对特定系统并具有明确的研究或设计等目的，从而一方面应当反映所关注的系统主要特征，同时又具有可操作性. 作为系统模型，首先必须界定系统边界，它的控制输入、干扰输入、状态和输出，以及它们之间的动态演化关系，其表现形式大多是一些定量或定性的数学关系式，但也可以是用于仿真和分析的计算机程序，或是某种具有共同特征的较简单的物理模型，以及它们之间的混合形式等. 系统建模常用的有两大类途径，以及它们的结合. 机理建模按系统的物理性质和问题需要选择适当的变量加以描述，并由已知的事物规律建立这些变量之间的关系. 数据建模则根据观测或实验数据，用各种设定的关系进行拟合. 实际上应用的往往是机理和数据的结合，即有规律可循时尽量用更能反映系统运动规律的机理模型，而对机理不明或过于复杂者则采用数据模型. 对于大型复杂系统，则常要用多种定性或定量的、处于不同层次上、不同质的多个子系统，综合集成地加以描述.

搜索技术(search technique) 用搜索方法寻求问题解答的技术. 具体常表现为系统设计或达到特定目的而寻找恰当或最优方案的各种系统化的方法. 当缺乏关于系统或这些参数的足够知识时，很难直接达到目的，诸如在博弈、定理证明、问题求解之类情形. 因此，搜索技术也是人工智能的一个重要内容. 当待搜索方案的集合(称为搜索空间)具有离散的树状结构的情形时，则可用启发式的规则来加快搜索过程. 常见的有深度优先极小化极大法和 α-β 法等. 又当待搜索的是一维或多维空间中的数值 x (标量或矢量)时，则可以把搜索目标定义为使某个品质函数 $f(x)$ 的值为最大. 常见的困难在于 $f(x)$ 的构造太复杂或者不十分了解，常规的数学分析的方法难以运用. 特别是当 $f(x)$ 具有多峰特性时，基于梯度驻点条件的方法都很难保证给出全局极大，而只能得到局部极大. 除了最简单的扫描搜索或盲目搜索外，有若干结构化的方法可以加速搜索过程，其中较重要的有斐波那契搜索、随机搜索等.

投入产出分析(input-output analysis) 亦称部门关联平衡法. 最初作为一种数量经济分析方法，于 20 世纪 30 年代由美国俄裔经济学家列昂节夫(Leontief，W.)提出，用以研究和分析各种经济活动的投入和产出之间的数量关系，特别是国民经济各部门在产品的生产、积累和消费之间的数量关系. 所谓投入，指从事该项经济活动的消耗. 例如，生产过程中所消耗的材料、能源、机器设备、劳动力等. 所谓“产出”，则指从事该项经济活动的成果. 例如，得到一定数量的产品、实现一定的运输量或完成某一种服务活动等. 投入产出分析的主要内容是建立投入产出表和投入产出模型，进行各部门经济指标分析，并在分析的基础上制定计划或发展规划. 投入产出表可分为价值型和实物型两大类. 价值型投入产出表是以货币为单位，它便于计算，便于各部门之间的比较；实物型投入产出表则不受价格因素的影响，但有时缺乏可比性. 投入产出表的一般形式如下表所示.

投入产出的数学模型，主要用于说明总产品和中间产品、最终产品之间的关系，以及总产值和物质消耗、净产值的平衡关系式. 通过模型分析，既体现经济系统的整体性，又可按部门分解以降低复杂性. 二者的有机结合还能从生产和分配两个方面反映产品在部门之间的运动过程. 通过各种系数反映技术水平、管理水平和部门间的技术经济联系.

当考虑时间因素时，还产生了动态投入产出分析. 从而，投入产出模型，按照分析时期不同，可分为

静态模型与动态模型两类；而按模型的编制范围不同，还可分为世界模型、多国模型、全国模型、地区模型、部门模型、企业模型等.

投入产出分配表

<table>
<tr><td rowspan="3" colspan="2">投 入
（产品的消耗来源）</td><td colspan="3">产出（产品的分配去向）</td></tr>
<tr><td>中间部门</td><td>最终产品</td><td rowspan="2">总产品</td></tr>
<tr><td>部门1 部门2 … 部门n 合计</td><td>累积 消费 合计</td></tr>
<tr><td>物质消耗</td><td>部门1
部门2
⋮
部门n
合 计</td><td>Ⅰ</td><td>Ⅱ</td><td></td></tr>
<tr><td>净产值</td><td>劳动报酬
纯收入
合 计</td><td>Ⅲ</td><td>Ⅳ</td><td></td></tr>
<tr><td></td><td>总产值</td><td></td><td></td><td></td></tr>
</table>

从投入产出分析的发展趋势来看，目前正与系统科学和经济计量学方法相结合，使之更为实用和有效.

马尔可夫决策过程(Markov decision process) 简称马氏决策过程. 一类无后效，即马氏过程的决策问题. 由于该过程的未来演化特性只依赖于其当前状态而与历史无关，马氏决策规则亦只考虑其与当前状态的依赖性，从而可使问题大为简化，并得到广泛的应用. 马氏决策理论最早由美国数学家贝尔曼(Bellman, R.)和霍华德(Howard, R. A.)分别于1957年和1960年研究并出版专著，其后在多步决策、最优控制和随机优化等方面引起了广泛注意，并对其进行了大量深入的研究和应用.

马氏决策过程的基本模型是一类随机递推演化过程. 以较简单的离散时间情形为例，其由第 k 步状态 $x(k)$ 向第 $k+1$ 步状态 $x(k+1)$ 转移的条件概率分布只与状态 $x(k)$ 及第 k 步的决策量 $u(k)$ 有关，而与此前的历史无关，可以写成

$$P[x(k+1)\mid x(k)] = f[x(k), u(k)]$$

($k=0,1,\cdots,N$；N 可能为无穷). 决策的目标通常为求 $u(k)$，使依赖于所有状态和控制作用的某品质函数 J 达到极大. 亦即

$$J = J[x(0), x(1), \cdots, x(N); u(0), u(1), \cdots, u(N-1)] \rightarrow \max.$$

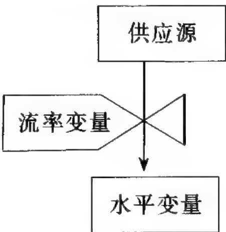
由于马氏决策的无后效性，它满足"最优性原理"，并可用动态规划方法求解(参见"动态规划"). 对于动态规划的主要困难"维数灾"问题，则又发展了许多迭代技术，如值函数递推方法、策略空间逼近等，近年根据实际问题需要又发展了 Q(价值)函数强化学习方法、随机逼近和扰动分析等用于处理高复杂度问题的近似方法，并对突发事件驱动的通信网络、生产、加工系统等排队网络模型的调度. 优化问题提供了实用的算法.

动态规划(dynamic programming) 求解多步最优决策过程的一种迭代方法. 由美国数学家贝尔曼(Bellman, R.)于1957年提出. 对无后效的马尔可夫多步决策过程，每一步的决策只依赖于当时状态及其后的系统演化，而与前面的历史无关，从而可知其后半段的轨道本身亦应是一条最优轨道. 这就是动态规划的基本依据"最优性原理". 由此即可推出基本的贝尔曼递推方程. 由于它把原来的多步决策问题嵌入一族递推的问题之中，这种方法亦被称为"不变嵌入原理". 动态规划也是解决离散时间或连续时间最优控制问题的一种有效方法.

动态规划虽可减少计算量，但它要把各步的最优控制结果全部记录下来，要求很大的存贮空间. 该存贮量依问题的维数而指数增长，被称为"维数灾"，已成为应用动态规划方法要解决的主要问题.

系统动力学(system dynamics) 研究反馈系统动态行为的一种基于计算机仿真的理论和方法. 由美国麻省理工学院福雷斯特(Forrester, J. W.)教授创立. 系统动力学用水平变量、流率、加、乘、积分和一些非线性环节连结而成，还可能包含若干反馈回路，用以模拟企业、城市、经济、社会、环境等各种复杂动态系统行为. 他提出了一套用于预测、决策和调整政策的分析方法，已形成了一套概念、原理、方法和称为 Dynamo 的程序语言和软件. 根据应用领域不同，福雷斯特等人已经出版了《工业动力学》《城市动力学》《世界动力学》，以及更概括的《系统动力学》等书. 国内外均已成立系统动力学研究和应用的协会和专门机构.

系统动力学模型本质上是带时滞的一阶微分方程组. 该方法建模时借助于流图概念，其中流位变量、流率变量、辅助变量等都具有明确的物理(经济)意义，是一种实用化的建模方法. 它的基本方法可归结为：确定系统状态的水平变量和流率变量；建立各影响因素的因果关系和流图，包括由流率到水平变量的积分单元及某些非线性环节，



它们可以用图形或方程组来表示；进行计算机仿真

和各种定性或定量的分析.图为其流率变量和水平变量之间最基本关系的表示法.

由于现代管理决策系统多被认为是一个开放的动态系统,因而上述组织系统框图中还应包括决策实施过程中的"反馈调整机制",使人们可以遵循一定的决策程序,不断进行检查调整,尽可能减少决策失误.

水平变量(level variable)　系统动力学中状态变量的一种.它表示某些物理量的累积水平,在结构图或方程式中它可由流率变量的积分来表达,并可由流率和以控制它与流率之间关系用图形表示,参见"系统动力学"条目之图.

流率变量(rate variable)　系统动力学中状态变量的一种.它表示另一类称为水平变量的状态变量的变化率,并可通过积分环节来计算后者.流率变量可以是自然确定的,或为人为施加的控制量.在绘制系统流图时,它和水平变量间之关系用图形表示,参见"系统动力学"条目之图.

反馈回路(feedback loop)　这里特指系统动力学中的概念.即指系统动力学中存在闭合因果链以描述系统诸要素间循环因果链的闭合回路.反馈回路有两类:一类是有自我抑制作用的信息反馈,称为负反馈回路;另一类为具有自我强化作用的信息反馈,称其为正反馈回路.

Dynamo **语言**(Dynamo language)　系统动力学的专用计算机语言.Dynamo 是英文 Dynamic 和 Model 的缩写.它是由系统动力学创始人福雷斯特(Forrester,J.W.)所在的美国麻省理工学院斯隆管理学院系统动力学研究小组创立的.

Dynamo 编译程序是遵循系统动力学使用的结构规则和运算规则建立的模型计算程序.它根据系统动力学的结构原理,整理模型,进行各类方程分类,安排那些相互依赖的辅助方程所需要的计算顺序,并依据模型编制的程序,把具有代数符号的方程转换成计算机操作指令,最后,根据所给定的解区间和运算时间长度的控制指令,迭代计算并输出系统仿真结果,从而它是一种动力系统建模和分析的通用工具.其后又相继开发出 Dynamo-Ⅱ 等更具一般性的新版本.

工业动力学(industrial dynamics)　系统动力学在工业企业管理中的应用.它综合运用控制工程、控制理论、组织理论和计算机模拟等方法,研究工业系统中的信息反馈特性,特别是各物理量的放大、叠加、计量积分、延迟和若干典型的非线性变换效应,及其对企业当前及未来的影响,从而为企业结构改革和策略制定提供依据.这一分支由美国麻省理工学院福雷斯特(Forrester,J.W.)于 1980 年出版的《工业动力学》一书而奠基.工业动力学讨论的模型和问题包括:生产和分配系统,后者包括库存、销售、广告和市场等环节;用户-生产者-雇佣系统,包括其周期演变特性及参数调节的灵敏度等问题;进一步的推广应用包括市场动力学、企业活动和生产的增长、各种产品的生命周期、研究开发部门管理、高层管理决策、现金和核算、竞争策略、人的作用和管理教育等.

世界动力学(world dynamics)　应用于全球问题的系统动力学.面临当代全球迅速发展所出现的问题,急需研究各种力量和行动相互作用而产生的总体后果.在美国麻省理工学院福雷斯特(Forrester,J.W.)于 1971 年出版的《世界动力学》一书中,研究了社会系统的结构,建立了人口回路、资本投资回路、环境污染回路等子系统模型.针对增长极限问题,书中讨论了自然资源减少、环境污染、人口爆炸、食物短缺等问题,并提出了若干补救措施.所用的建模方法,包括在大脑中建立起的概念模型和可用数据加以实现的计算机模型等,然后用 Dynamo 软件进行仿真.作者承认该书的模型还是十分简化和粗糙的,20 世纪 70 年代以来在这方面又有许多新的成果.

最优搜索问题(optimal search problem)　一种追求成功概率最大的随机搜索问题.它研究如何在总费用受约束的条件下,使探测到目标的概率达到极大.其形式定义如下:设被搜索目标为固定,它可能处于某有限或可数的正整数集 J,并具有离散的概率分布 $p(j)$,或在欧氏空间 X 中具有连续概率分布密度 $p(x)$.探测器沿某曲线运动时,它探测到目标的概率 $\hat{a}(r)$是它与目标之间最短距离 r 的函数.搜索的代价用搜索力来度量,它可以是探测器的行程距离或搜索时间等.搜索问题归结为确定施加给搜索空间 J 或 X 的每一单元的搜索力的分配函数 $f: J\to\mathrm{R}^+$ 或 $X\to\mathrm{R}^+$.而探测函数 $b(j,z)$定义为把 z 数量的搜索力施加于 J 或 X 中的单元 j 而在 j 处探测到目标的概率.从而对应分配函数 f 和目标分布密度 $p(j)$或 $p(x)$探测到目标的总概率为

$$P[f]=\sum_{j\in J}p(j)b[j,f(j)]$$

或

$$P[f]=\int_X p(x)b[x,f(x)]\mathrm{d}x.$$

费用函数 $c(j,z)$则定义为把搜索力 z 施加于单元 j 的费用,从而对应每一分配函数 f 可以求出其总费用为

$$C[f]=\sum_{j\in J}c[j,f(j)]\ \text{或}\ \int_X c[x,f(x)]\mathrm{d}x.$$

于是最优搜索问题可形式地定义为在约束 $C[f]\leqslant K$ 之下求 $f^*\in F$,使

$$P[f^*] = \max_f P[f].$$

搜索问题的研究始于第二次世界大战期间对敌方潜艇或飞机的探测要求，其后曾用于1966年搜索丢失于地中海中的氢弹，1968年搜索失事的核潜艇斯科平(Scorpion)号，1974年清除苏伊士运河中的水雷等重大行动．已有的理论包括前述的对静止目标的最优搜索问题的求解，对有假目标存在时的搜索、最优搜索与停止，对若干运动目标的搜索等.

斐波那契搜索(Fibonacci search)　一种有限区间中单峰函数的搜索技术．为简单起见，设此区间为 $L_1=[0,1]$，记 $\{F_i\}$ 为斐波那契数，

$$F_n = F_{n-1} + F_{n-2},\ F_0 = F_1 = 1.$$

第一次估值点为

$$x_1 = \Delta_2 \triangleq \frac{F_{n-2}}{F_n} \quad 和 \quad x_2 = 1 - \Delta_2,$$

其中，$1/F_n$ 应等于或小于搜索的预期精度．若 $f(x_1) > f(x_2)$，则删去 $(x_2,1]$，反之删去 $[0,x_1)$．以 L_1 记删去的区间，再对留下的区间 L_2 取

$$\Delta_3 = l_2 \frac{F_{n-3}}{F_{n-1}}(其中 l_2 为 L_2 之长度),$$

对 L_2 重复上述步骤．如此反复直到

$$l_n = l_1\left(\frac{F_0}{F_n}\right).$$

已经证明，斐波那契搜索是一种函数估值次数最少的最优搜索方法.

随机搜索(stochastic search)　一类将量度点随机化的多维空间搜索技术．它包含以下几种方法：

1. 真随机搜索．将搜索区域分为有限 N 个多维单元，假定最佳的 n 个单元均满足搜索目标的要求，随机任取一单元，即满足此要求的概率为 n/N，进行 m 次这种取样，则至少能取得一满意结果的概率为 $1-(n/N)^m$．这样可用较少的次数选到一个满意的结果.

2. 蠕行随机搜索．先由验前知识选择一预测点，并按以它为中心的一种非均匀概率分布密度进行随机搜索．再由取得的若干后继取样点中选取一最佳者作为新的中心，并重复类似的搜索，直到满足搜索目标的要求.

3. 随机搜索．按上述方法选定预测点，以此为中心进行一次搜索后，需将该试探点与原中心进行比较，若有改进则将中心移到该新试探点，并再进行非均匀分布的搜索，直到进行若干次试探均未能改进时停止搜索.

4. 加速随机搜索．按上述方法选预测点进行搜索至发现改进点方向后，将新的中心放在改进点并向前延伸一步，继续这一过程直到无可改进为止．比较起来，这种方法收敛速度最快，而且不仅适用于单峰，也适用于多峰搜索的情形.

贝叶斯决策(Bayes decision)　一种具有学习能力的随机决策方法．设系统状态为 Θ，其验前分布密度为 $\pi(\Theta)$，实际可得的观测值为 X，设已知观测值 X 以 θ 为条件的分布密度(亦称似然函数)为 $f(x|\theta)$，又设由观测值 X 确定决策变量的规则为 $\delta(x)$，对应损失函数为 $l(\Theta,\delta(X))$、风险函数为 $R(\Theta,\delta)$，并进而可求出贝叶斯风险 $r(\pi,\delta)$，则贝叶斯决策问题意指为求决策规则 δ 使贝叶斯风险 r 为最小.

贝叶斯决策问题的主要特点是在验前知识不完全时，可以利用实验和观测到的验后数据来提高已有知识，从而具有某种学习功能．验后密度 $\pi(\theta|x)$ 和验前密度 $\pi(\theta)$ 及条件密度 $f(x|\theta)$ 之间的关系，即为著名的贝叶斯公式

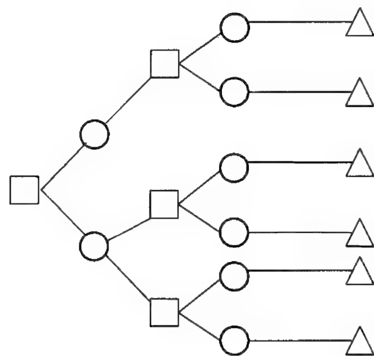
$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)},$$

其中归一因子

$$m(x) = \int_{\theta\in\Theta} f(x|\theta)\pi(\theta)\mathrm{d}\theta,$$

称为预测密度.

决策树(decision tree)　一种描述决策因素相互依赖关系的树形图．图示为其一例，其中，□表示"决策点"(又称节点)，从它引出的分岔称"方案分岔"，各代表一个行动方案节点；○表示"自然状态"，其上标明损益期望值，从它引出的分岔称"概率分岔"，表示自然状态及其出现概率；△为"结果节点"，其数字表示每一方案在其相应状态下的损益值．决策树用以形象地描述风险决策，由之可一目了然地标明各种更换替代方案、各种可能出现的状态及可能性大小及产生的后果，从而为集体决策提供了很大方便.



冲突分析(conflict analysis)　专门研究和解决现实世界冲突现象的学科分支．它由经典博弈论中的超对策论发展而来，具有严格的数学基础；同时，又避免了传统博弈论的两个基本假设，即效用假设和传递假设，从而成为一种更为实用的处理现实世界中冲突问题的方法．冲突分析的过程分为两步：第一步是建模；第二步是根据模型和所获得的信息确

定可能的决策方案，即求出对所有局中人来说都是稳定的结果. 稳定是指对局中人，仅通过改变自身策略已不能得到任何改进的结果. 冲突分析由加拿大的海普尔(Hipel, K. W.)等于 1979 年提出，曾用来分析如古巴导弹危机、多国用水争端、跨国公司权益等冲突问题，形成了一套工作步骤和计算机软件包.

聚类分析(cluster analysis) 一种对事物和数据进行分类的方法. 即在不具验前类别信息情形下，单纯由样本数据集合根据某种相似性度量对事物进行分类的方法. 通常先定义各模式特征向量之间的距离，把距离较小的样本归为一类，再检测类间的距离，看类别是否显著. 常见的聚类分析方法有谱系聚类法(系统聚类法)和动态聚类法. 前者每次将距离最近的两类(或个体)合并，再计算新的类间距离. 这样每次减少一类，直到达到某种满意指标为止. 后者则先设定类别数，各类均有一初始代表样本(类中心). 按距离把各样本划归离中心最近的一类，再根据新形成的类计算新的中心，直到不再变化. 之后还要根据类间的接近度和类内的散布度对有关的类进行合并或分离，直至满意. 聚类分析适用于验前知识较少的测量数据处理和自然规律探索，实际上是一种无监督的学习方法.

计划评审技术(program evaluation and review technique) 简称 PERT. 一种管理统筹方法. 由美国海军部一个由数学家和工程师组成的研究小组于 1958 年提出. 计划评审技术是一种基于工序流程图的作业调度优化技术，它和关键路径法的主要区别在于：后者是以经验为基础来确定工序时间，而计划评审技术则适用于没有经验的新项目，它对各项活动的时间估计是根据最长、最短、最可能时间，并考虑概率因素确定的. 计划评审技术注重各项任务安排的评价和审查，并因此而得名.

关键路径法(critical path method) 简称 CPM. 一种计划管理统筹方法. 1956 年，由美国一个包含数学家和工程师的研究小组提出. 关键路径法的基础是工程项目的工序流程图，它由许多工序(活动)组成，每个工序用一个箭头代表，工序完成所需时间称工序的长度，用时间 t 表示. 工序的交接处画一个圆圈，表示工序的开工和完工事项. 这样就可以把整个工程按先后顺序，从左至右，按逻辑顺序从开始到结束用网络形式的工序流程图表现出来. 由此图再画出路径，即从起点顺箭头方向连续不断地到达终点的一条通道. 路径的长度定义为其上各工序时间长度之和. 关键路径指的就是最长的路径. 显然，关键路径指出了工程的关键性所在. 只要在网络图中加强和优化关键路径，如将非关键路径上的人力和设备往关键路径上调动，就可以达到缩短工期、提高工效、降低成本的目的. 关键路径法多用于确定性问题的研究.

时间序列预测技术(time series forecasting technique) 依据时间序列进行预测的方法. 所谓时间序列指的是通过观测所得到的离散时间观测点上的系列数据，通常可用 $y_1, y_2, \cdots, y_t$ 或 $\{y_t\}$ 表示. 时间序列预测技术根据事物演化先后状态间的相关联系可以来预测该事物的未来发展. 具体方法主要有移动平均法、指数平滑法和季节(周期)性模型等. 移动平均法是对一时间序列取其最近若干个数据，求取其加权平均值，作为对未来的预测值. 为了避免其上升或下降趋势带来的偏离，可以用二阶移动平均. 亦即利用一次移动平均法所得到的数据序列，再进行一次移动平均计算. 类似还可推广至三阶、四阶，乃至高阶的移动平均. 指数平滑法引入指数式的平滑系数进行移动，平滑系数反映其对历史数据重要性的修正，取值于[0,1]之间. 一次指数平滑法一般适用于水平型数据，类似地也有二阶指数平滑法等. 此外，还有季节(周期)性模型，又称为 Winter 模型，适用于对具有周期性的数据序列进行预测，它考虑了季节性周期长度及水平. 采用周期性、趋势性平滑常数，可以较好地对周期性时间序列进行预测.

白箱(white box) 系统科学术语. 对系统内部机制基本清楚的一种形象说法. 通过对系统的观测，对于所研究对象的系统结构、运动过程、动态特性有比较全面认识的系统和部件.

黑箱(black box) 系统科学术语. 对系统内部机制基本不明的一种形象说法. 由于对象的系统结构、运动过程、动态机制难以被人们所观测，人们只能凭系统的输入、输出关系来推断系统的结构和部件.

灰箱(grey box) 系统科学术语. 对系统内部机制知之不全的一种形象说法. 它介于白箱、黑箱概念之间，指的是对所研究系统内部结构、运动过程、动态特性可以部分地观测和认识，而不能完全了解. 这时将两方面结合起来，有可能得到较为完整的认识和有效的控制.

自回归滑动平均模型(autoregressive moving average model) 简称 ARMA 模型. 平稳随机过程序列预测方法的一种. 可以看出 ARMA 模型是 AR 模型(自回归模型)和 MA 模型(滑动平均模型)的综合. 其一般形式为

$$z_t = \varphi_1 z_{t-1} + \cdots + \varphi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q},$$

其中，z_t 为时间序列变量，a_t 为随机干扰或控制作用. 应用这种模型进行预测的步骤是：

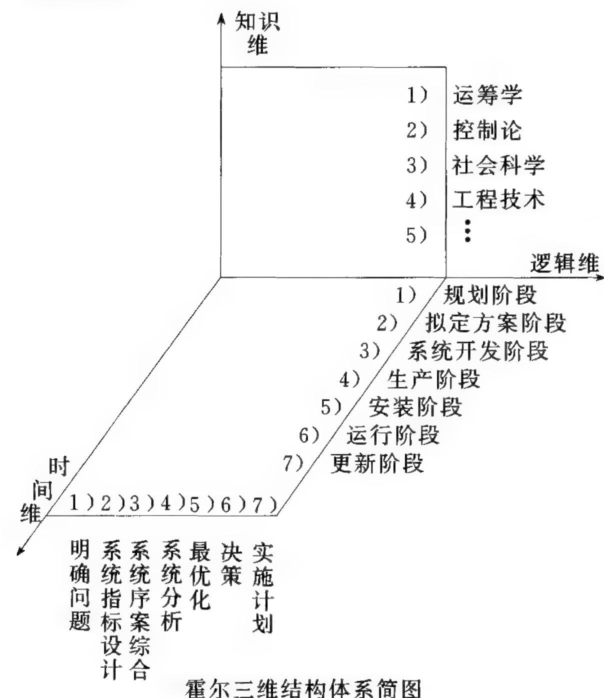
1. 模型识别. 即判断问题是否适于采用 ARMA 模型.

2. 参数估计. 根据识别模型及其阶数，对有关参数进行估计.

3. 模型检验. 即在模型识别和参数估计的基础上,用统计检验方法,检验模型是否合理、准确.

4. 模型预测. 即应用建立的模型对系统的未来进行预测. 对于 ARMA 模型来说,在模型识别和参数估计过程中,既要用到 AR 模型的概念,建立辅助序列,计算自相关函数,又要构成 MA 模型,估计其参数.

霍尔三维结构(Hall three-dimension structure) 展示系统工程概念和方法的一种结构图. 1969 年,由美国系统工程专家霍尔(Hall,A. D.)提出. 它由时间维、逻辑维和知识维构成,用以展示系统工程各项工作内容、步骤和相关知识. 它集中体现了系统工程方法的总体化、综合化、最优化、程序化和标准化的特点,概括地表示出系统工程各个阶段、各个步骤以及所涉及的知识范围,是系统工程方法论的基础.



霍尔三维结构体系简图

1. 时间维. 反映具体的工作进程. 一个具体工作项目可以分为七个阶段,即:

1) 规划阶段.

2) 拟定方案阶段.

3) 系统开发阶段,系统研制方案及详细计划.

4) 生产阶段,零部件生产,拟定安装计划.

5) 安装阶段,系统安装,提出运行计划.

6) 运行阶段,投入运行,展开服务.

7) 更新阶段,改进旧系统,建立新系统.

2. 逻辑维. 描述解决问题的逻辑过程. 运用系统工程方法解决大型项目,一般要经历七个步骤:

1) 明确问题. 通过调查,明确所提出的问题及其确切要求,尽量全面收集有关问题的历史、现状和发展趋势等方面的资料.

2) 系统指标设计. 选择系统功能评价指标,设计评价算法,组成评价指标体系.

3) 系统方案综合. 按问题性质和总体功能要求,综合形成一组备选系统方案,明确其结构和参数.

4) 系统分析. 分析系统各方案的性能、特点及其实现程序,按评价目标体系确定其优劣次序.

5) 最优化. 在约束条件下,综合比较各备选方案,选出最佳方案.

6) 决策. 在分析、评价和优化基础上做出裁决,选定行动方案.

7) 实施计划. 根据选定的方案,将系统付诸实施.

这七个步骤只是一个大致的过程,其前后次序有时也不严格,而且往往要反复多次才能达到最后的目标.

3. 知识维. 描述完成各个阶段和步骤所需的各种专业知识、技能和技术素养.

霍尔三维结构有助于理解多级递阶控制法、结构模型解释法、计划协调技术和图解协调技术等方法的应用,以及大型项目总体模型的建立. 它还强调程序、人和工具三者的精心协调,并突出了人在其中的主导作用.

多学科团队决策模型(interdiscipline team decision model) 一种多人决策模型. 它包括来自不同学科的决策群体按共同目标进行决策的工作模式及其规范的模型. 从事大规模复杂人机系统的分析或设计的组织,其核心必须具有由多学科成员组成的团队决策行为的模式. 按照维默(Wymore,A. W.)的定义,这种团队应包含若干与其面临问题有关的不同方面受有良好训练的成员,协作处理系统的大规模和复杂性、动态演化问题和技术、以及人类行为的建模和分析等问题. 其典型例子之一是现代企业产品开发的并行工程,它将产品开发的各个阶段,诸如创意、初始设计、原型设计、工艺设计等方面的人员通过多学科团队协作而同时进行,它运用现代信息技术使设计过程中的问题随时得到解决,从而可成倍地缩短开发时间,加速了国防和经济建设的步伐,提高企业的竞争力. 该原理同时也可用于其他领域. 维默还具体建立了一种系统设计的"三子叶"理论模型来描述多学科团队的决策过程(参见"系统设计三子叶理论").

系统设计三子叶理论(tricotyledon theory of system design) 系统设计过程的一种数学模型和理论. 由维默(Wymore,A. W.)提出. 它也是支持各不同阶段多学科团队决策的一种系统工程方法论.

三子叶理论把系统设计问题划分为:

1. 输入/输出子叶. 它规定了系统的输入/输出关系,可以用各种数学关系,也可直接用表格等形式.

2. 技术手段子叶. 它规定了所设计系统可以采用的解决问题的技术手段,常用一种“工具箱”的形式来表示.

3. 可行性子叶. 它是既符合系统输入/输出要求,又可用已有技术来实现的部分,实际上也就是前两个子叶的交集.

层次分析法(analytic hierarchy process) 简称 AHP. 多目标决策问题的一种分析方法. 又译为“解析递阶过程”,由美国运筹学家萨蒂(Saaty,T. L.)于 20 世纪 70 年代提出. 该方法将有关决策问题的因素分解成目标、准则、方案等层次,而后在此基础上运用定性与定量分析相结合的决策方法. 基于对决策问题的内在本质因素及其关系进行的深入分析,利用较少的定量信息把决策者的经验和思路数学化,从而为多目标、多准则或无结构特性的复杂问题提出的一种简便可行的决策方法. 这种决策方法和决策问题通常采取以下步骤:

1. 把复杂问题分解为各组成元素,建立一个多层次的递阶结构,以确定决策问题各元素之间的递阶关系. 首先分析决策问题所包含的元素及其相互关系,根据目标的要求和这些关系将元素分析成不同的层次. 对于元素少、元素间关系较明确的决策问题,可以直接凭经验建立结构,否则要采用解释结构模型的方法来建立结构.

2. 建立判断矩阵,据以计算各元素的优先级权重. 用上一层次的各元素作为下一层元素的判断标准,分别对下一层元素两两比较,比较其对于标准的重要程度,同时定量化,建立判断矩阵. 计算判断矩阵的最大特征值及其正交化特征向量,得出这层元素的优先级权重.

3. 确定决策问题的总体优先级的组合权重. 把上一层各元素作为下一层元素的判断准则,得出下层元素对上层各元素的权重,然后用上层元素的组合权重加权平均,得出下层各元素的组合权重,并用于决定下层元素的组合权重. 这种方法可以得到各层元素对于总体目标的组合权重.

4. 分析计算结果,做出相应决策.

分解(decomposition) 大系统简化分析的一种方法. 通过分析大系统的各子系统,然后加以综合,以降低信息处理和控制的复杂程度的一种方法. 所谓子系统指大系统中相对独立的部分,其内部耦合较紧而与大系统其他部分关联比较松弛. 由于系统分析的复杂度通常至少为其维度的高次多项式,这种分解可大为简化大规模系统分析的工作量.

集结(aggregation) 大系统简化分析的一种方法. 将大系统关系较为紧密的许多状态变量或单元归并,从而化为较少的集结状态变量或组合单元,以求降低系统分析的复杂度. 集结法是大系统建模和模型简化的重要方法之一. 常见的是通过线性组合把众多状态变量归并成少数新的状态变量. 集结法于 1948 年在建构宏观经济模型时提出,其后许多人对此又进行了深入的定量研究.

设大系统模型为

$$\dot{x} = Ax + Bu, \quad x(t_0) = x_0, \tag{1}$$

上式中,x 为 n 维状态向量,u 为 r 维控制向量,A 和 B 是具有相应维数的矩阵. 用 $m \times n$ 阶集结矩阵 $C(m<n)$,按 $z=Cx$ 将高维状态向量 x 集结为低维状态 z. 易证这时 y 满足方程

$$\dot{z} = Fz + Gu, \quad z(t_0) = Cx_0, \tag{2}$$

式中 F 为一个 $m \times m$ 矩阵,G 为 $m \times r$ 矩阵,它们显然应满足必要条件

$$FC = CA, \tag{3}$$

$$G = CB. \tag{4}$$

但若要求(2)是(1)的一个完全集结的简化模型,则还需满足一些兼容性条件. 由于矩阵 F 的特征值是原方程中 A 的特征值的子集,如果 F 的 m 个特征值是(1)的主导特征值,即被略去的 $n-m$ 个特征值对应快衰减振型,则(2)的动态特性与(1)的动态特性差别甚微. 条件(4)保证稳定态时集结关系 $z=Cx$ 成立. 简化模型(2)导出的反馈控制作用于原系统时,仍可保持其稳定性和品质. 由于满足(3)的 F 并不总存在,所以一般只能得到近似集结的简化模型. 简化模型阶次 m 的选取,与系统非主导快衰减振型的数目有关.

模型降阶(model order-reduction) 简化处理大规模系统的一种方法. 为用低阶模型近似代替原有的高阶模型,以简化分析、计算并保持原系统的主要性质. 主要方法包括用线性变换以低阶状态向量 Fx 代替原高阶变量 x 的集结法,在状态空间模型中保持主振型的方法,对传递函数模型用 Pade 逼近或连分式进行降阶的方法等.

大(规模)系统(large-scale system) 规模庞大、结构复杂的系统. 其庞大、复杂程度的具体界定需由问题、知识、解题手段等共同确定. 目前则主要以其基于递阶层次结构和子系统分解、分布式信息处理和控制的基本思想和方法论来加以定性. 这类系统的主要特征可以概括为:

1. 构成要素种类多、数量大,其状态变量维数高到难以集中处理的程度.

2. 系统中信息是分散、分布式、多时间尺度的,且信息量大至难以完全采集和处理的程度.

3. 子系统数量多,时空变化范围大,并具多层次和分散的控制结构.

4.各子系统控制和优化的目标是多种、多层,甚至互相冲突的,它们常代表其中不同人群的利益和价值标准,从而导致多人决策和博弈论方法的应用.

大系统理论是20世纪70年代人们将现代控制理论推广应用于社会经济、企业管理、人口资源、生态环境、电力系统以及通信、交通等更大范围问题时提出的,当时主要关注由于时空规模大,变量及层次数量多带来的困难.其主要内容包括:大系统建模所用的分解、集结、奇异摄动等降阶简化技术;大系统稳定性判别和镇定;递阶和分散控制理论;随机分散控制和信息结构问题;多层次、多人、多目标协调优化和控制理论等;以及它们在各方面的应用.20世纪80年代以后,随着耗散结构、协同学、突变论等复杂系统理论的出现,人们又逐渐转向复杂系统和复杂性的研究,更强调系统宏观、微观之间质的差异.

协调(coordination) 大系统分解后处理各系统间关联的一种手段.由于所处理系统的庞大,为降低信息处理的复杂性,可采用分解/协调的概念和优化算法.先把大系统分解为若干子系统,分别求取子系统的最优解,再考虑子系统之间存在关联,进行子系统之间的协调,反复迭代最终逼近大系统的最优解.在分解/协调算法中,有目标协调法,或称非现实法、不可行法,另外还有约束协调法,或称现实法、可行法,以及综合这两者的混合协调法.大系统优化过程中,可能还需要不断进行上下级之间的关联,也称为协调.

系统工程方法论(system engineering methodology) 关于系统工程研究方法的理论.它具体包括为解决系统工程问题或实施系统工程项目所需的一套概念、模型、分析手段、设计和实施过程乃至运行管理的工具、方法和解题步骤.由于系统工程面临的多为复杂系统的问题,通常单一学科的概念、知识和方法不足以解决这些复杂问题,方法论是必须的.系统工程发展过程中,许多国内外专家都十分重视系统工程方法论的研究,并出版大量专著.随着系统开放性和复杂性问题日益突出,中国的钱学森等科学家又提出了开放的复杂巨系统从定性到定量综合集成的方法论.

德尔菲法(Delphi method) 一种专家意见征询方法.由美国兰德公司提出.其步骤为:

1.根据任务归纳出含义明确、具有统一量化评价标准的问题.

2.选择适量的、具有代表性的专家,并逐一分别征询各位专家的意见.

3.回收专家意见,按前面所述的统一量化评价标准进行统计分析,形成对问题看法的初步统计结果.

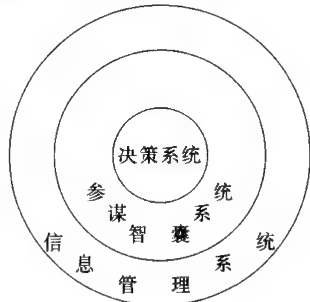
4.将初步统计结果再反馈给上述范围的专家,再次征询专家意见,这一次专家可修改自己前次的看法.

5.经多次反复,形成最终意见.

德尔菲法的优点是对问题评价过程中,比较少地受权威意见干扰.由于可向多方面的、多位专家、多次反复征询意见,具有较高的可信性.其不足之处在于,缺乏专家看法之间的启发及不同意见的协商.

决策分析(decision analysis) 一门新兴交叉学科.它是运用现代决策、控制理论和信息处理技术解决复杂管理、决策问题的理论和技术.决策分析作为一门学科是20世纪50年代之后才兴起的.经过多年发展,许多学者应用经济学、社会心理学、组织理论、系统分析、运筹学、计算技术等,对决策分析的行为、过程、信息、方法、组织等各个方面进行了广泛而深入的探索和研究,形成了一系列理论和方法.它同有关学科,如运筹学、组织行为学、统计学、信息学、系统科学等交错在一起,具有新兴学科发展的共同特点.

决策分析过程大体上可以分为三个大步骤,即明确问题所在,提出和确定决策的目标;发现、探索和拟定各种可能的行动方案;分析对比,从各种可能方案中选出最合适的方案.也有些学者把这三个步骤分别称为参谋活动、设计活动和选择活动.这三个基本步骤是任何完整的决策过程所必不可少的.决策分析作为一门新兴的交叉学科,具有鲜明的方法论特点.这类科学的理论与方法往往相互交织在一起,很难有明确的界限而截然分开.在决策分析中经常使用的一些方法有系统分析方法、费用-效益分析方法、价值分析方法、敏感性分析方法、模型模拟方法、可行性研究方法、定性关系分析方法、战略目标法、指令性约束及其他社会经济环境分析方法等.



决策分析的组织指的是决策机构的体制问题.现代化的决策体制是由信息管理系统、参谋智囊系统和决策系统组成的统一体.如图所示.其中心是决策系统,围绕它的是参谋智囊系统,最外一层是信息管理系统.一般地,决策者不直接参与信息系统的管理工作,但参谋智囊系统向决策者提供的决策分析是以大量信息为依据的.作为一个决策者,为了更有效地决策,应该掌握某些重要的第一手资料,要亲自了解一些关键信息.作为信息管理系统,应向决策者及参谋智囊人员负责,在收集、提供信息时,应做到及时、准确和适用.

目标协调(objective coordination) 亦称关联

平衡法. 大系统递阶控制的一种方法. 即选择关联拉格朗日乘子 λ 作为协调变量的递阶控制方法. 它一般用于求解下列极值问题:

$$\max_{\lambda}\{\max_{\rho}\min_{x,u,z}L\} = \max_{\lambda}\left\{\sum \max_{\rho_i}\min_{x_i,u_i,z_i}L_i\right\}.$$

也就是在第一级,按来自第二级的预估协调变量 λ,求 N 个子系统中拉格朗日函数 L_i 的极值解. 在第二级,根据第一级输送的状态变量 x 和关联输入变量 z,通过求拉格朗日对偶函数 φ 的极大解来更新 λ 的值,然后再进入下一次迭代. 采用这种上、下级之间信息的迭代变换,直到关联平衡时为止. 这种算法是不可分解的,因为在迭代过程中关联方程不成立,其中间结果均不可物理实现.

约束协调(constraint coordination) 大系统递阶控制的一种方法. 在复杂大系统演化过程中,各个原来独立分散的元素或子系统将不断地相互作用,逐渐形成一个有机整体. 它们一方面具有各自的功能以实现共同的目标和利益,同时又为了其自身的利益,又要争夺所需的物质、能量和信息等资源. 从全局的利益对这种竞争进行协调的方法之一是使各元素或子系统受到系统整体和其他子系统的某种"约束". 这些约束的具体形式可为各子系统的边界条件、环境参数、资源分配以及用拉格朗日乘子表现的价格或惩罚项等. 通过设计和调控这些约束手段,可以较简单地实现对子系统的协调,以达到全局的目标.

影子价格(shadow price) 亦称客观制约估价、最优计算价格. 将线性规划应用于经济问题的一个概念. 在资源最佳利用条件下,生产资料增加每个单位所带来的收益. 当用线性规划求解资源最佳利用的收益时,影子价格就是各约束条件右端项对目标函数的边际贡献. 设目标函数 $Z=C_BB^{-1}b$ 和检验数 $C_N-C_BB^{-1}N$ 中都有乘子 $Y=C_BB^{-1}$. 若 B 是

$$\{\max Z = Cx \mid Ax \leqslant b,\ x \geqslant 0\}$$

的最优基,则 $Z^*=C_BB^{-1}b=Y^*b$,由此得

$$\frac{\partial Z^*}{\partial b} = C_BB^{-1} = Y^*.$$

所以变量 y_i^* 的经济意义是在其他条件不变情况下,单位资源变化所引起的目标函数的最优值的变化. y_i 的值代表对第 i 种资源的估价,这种估价是依具体单位的特定产品而存在的一种特殊价格,即"影子价格". 影子价格由美籍荷兰经济学家丁伯根(Tinbergen,J.)于 20 世纪 30 年代首先提出. 美籍经济学家萨缪尔森(Samuelson,P. A.)把影子价格称为商品的边际成本. 在苏联及东欧,影子价格被称为最优计划价格,由苏联学者康托罗维奇(Канторович,Л. В.)首先提出,它反映了稀缺资源的边际机会成本,是资源配置决策和计划工作的重要依据. 在费用效益分析中,影子价格并非由市场形成或实际存在,而是按照一定原则人为推算出的物品和服务的价格. 它可代替"失真"的实际价格,以防盲目简单地挪用. 统一测定影子价格的适用范围主要是:项目建议书阶段,项目的投入物和非主要产出物;项目可行性研究及评估阶段,项目的非主要投入物和非主要产出物. 统一测定的影子价格通常均以绝对数值或换算系数的形式给出.

撰 稿 王 翎 朱松春 刘式达 刘华杰 刘家业 江中岳 孙之荣 杨家本 谷 可 张维民 陈秉正 苗东升 郑应平 法京怀 欧阳弈孺 赵松年 姜 璐 黄 昫 蒋昌俊 谭 民

审 阅 于景元 朱照宣 孙增新 李培信 张天平 陈 禹 秦化淑 涂奉生 霍 伟

控 制 理 论

控制理论(control theory) 一门有广泛应用和重大实用价值的理论.是研究在一定限制条件下如何控制系统以实现预期目标的一门理论.控制是用来影响动态系统行为的一种手段,对系统的控制是根据对其输出的实际观测所做出的决策再由输入的作用来实现的,目的是使系统实现稳定和预期目标的运动.控制理论属于技术科学范畴,可区分为基础研究和应用研究两个层次.在基础研究层次中,强调采用精确的数学语言来阐述控制的特性和原理,最终从数学上建立分析和设计控制系统的理论和方法.在应用研究的层次中,着重于建立控制系统的工程化分析和设计方法,将控制规律转化为可在计算机上实现的控制算法和相应的软件,以及针对实际控制系统的具体特点采用控制理论的概念和方法解决实际问题.

控制理论的最早研究可追溯到1932年奈奎斯特(Nyquist,H.)对反馈放大器稳定性的分析.他在其著名的论文中提出了一种以系统频率响应形式表达的判据,可用于判定反馈控制系统是否稳定,以及如何通过修正系统的特性使反馈系统由不稳定变为稳定或改善其稳定度.1940年,伯德(Bode,H.W.)进一步引入了可使频率响应特性的运算和作图简化的对数增益图和线性相位图.1948年,埃文斯(Evans,W.R.)提出了根轨迹法,使以复变量理论为基础的分析和设计反馈控制系统的方法更趋成熟.与此同时,维纳(Wiener,N.)和柯尔莫哥洛夫(Колмогоров,А.Н.)于1940年分别提出了基于频率域分析的信号处理的滤波和预测方法,它和建立在奈奎斯特稳定判据与埃文斯根轨迹法基础上的控制理论一起形成了经典控制理论.1948年,维纳尝试把以工程问题为背景的经典控制理论的概念、原理和方法扩展到研究生物机理和神经系统,他的专著《控制论,或关于动物和机器中控制与通信的科学》标志着一门独立学科"控制论"的诞生.1954年,钱学森的专著《工程控制论》问世,成为控制论发展的又一个里程碑.经典控制理论的主要研究内容包括:单输入-单输出线性定常系统的预测、滤波和估计的原理和方法;单输入-单输出线性定常系统的稳定性、调节和补偿的原理和方法;特殊单输入-单输出线性系统的快速控制的原理和实现方法等.

20世纪50年代,贝尔曼(Bellman,R.)在研究动态规划中提出了"最优性原理——最优过程的任何最后一段都是最优的",成为最优过程应当满足的必要条件.紧接着,庞特里亚金(Понтрягин,Л.С.)等人提出了作为最优控制问题基本原理的"最大值原理".最大值原理从另一角度建立了最优控制应满足的必要条件,并从理论上完全解决了线性多变量系统的快速控制问题.20世纪60年代初,卡尔曼(Kalman,R.E.)将状态空间方法系统地引入到控制系统的研究中,提出了揭示控制系统结构特性的两个基本概念"能控性"和"能观测性".卡尔曼并和布西(Bucy,R.S.)一道,建立了基于时间域分析的新的预测、估计和滤波理论——卡尔曼滤波理论.这些概念、方法和理论的出现,标志着控制理论进入了现代控制理论的新阶段.随着现代控制理论的发展,相继形成了多个独立的分支,如系统建模与系统辨识、线性系统理论、最优控制理论、鲁棒控制理论、分布参数系统理论、微分对策、适应控制以及非线性控制系统理论等.

控制理论是在重大工程问题的需求推动下形成和发展起来的.经典控制理论的形成和发展受到了通信技术和火炮技术的深刻影响,现代控制理论则是在远程火箭技术与航天技术的推动下并受到电子计算机技术的促进而发展形成的.现今,现代控制理论的方法、概念和原理正在日益渗透到非工程系统领域,诸如社会经济系统、生物和生态环境系统、现代管理系统以及集工程和管理为一体的更为复杂的系统.这类复杂系统的基本特征,一方面表现为系统本身所在环境和运行机制的不完全确知性,另一方面表现在运行过程中总是包含有人的参与.对于这类复杂系统,目前已经发展了一些有效的控制方式,如分散控制方式,把计算机作为一个环节直接进入系统运行过程的控制方式,以及人的经验和行为直接参与系统运行过程的控制方式等.但是,针对这类复杂系统的更为一般的分析、控制和综合理论还远未形成.

经典控制理论

经典控制理论(classical control theory) 一种控制理论.这里指20世纪30年代至50年代期间形成的自动控制系统分析和设计的方法及相关的理论.其所研究的受控对象通常以时间上连续的线性常微分方程描述为主,一般只有一个控制变量和一个输出变量,因此又称单输入-单输出(SISO)系统,或单变量系统.对单变量系统的分析一般利用复变

函数论中的拉普拉斯(Laplace,P.-S.)积分变换,得到频率域中的输入-输出关系,又称传递函数.经典控制理论中所要求解决的核心问题归结为设计控制器,使系统能有一个合适的传递函数,或使被控变量的动态过程满足工程要求,首先是稳定,其他有过渡过程时间、超调量、振荡次数、偏差大小和控制功率(能量)的大小等.随着计算机的普及,系统中物理量的测量和数据处理采用了间断采样的方法.这时系统描述由常微分方程转化为差分方程,拉普拉斯积分变换改为采用 Z 变换,传递函数改为采用脉冲传递函数.其他诸如受控系统的性质及系统分析、设计的基本要求,都没有显著的变化.这类采样系统,虽然是在 20 世纪 50 年代或其后发展起来的,通常也被归入经典控制理论的范畴.

传递函数的极点(poles of transfer function) 经典控制理论的基本概念.使系统的传递函数取无穷大的变量(频率变量)值,该值称为传递函数的极点.对于由线性定常微分方程式表示的系统,传递函数定义为输出量和输入量的拉普拉斯变换式 $Y(s)$ 和 $U(s)$ 之比,它为一实有理分式函数

$$W(s)=\frac{Y(s)}{U(s)}=\frac{Q(s)}{P(s)}$$
$$=\frac{b_0+b_1s+b_2s^2+\cdots+b_ms^m}{a_0+a_1s+a_2s^2+\cdots+a_ns^n},$$

其中 $Q(s)$ 和 $P(s)$ 分别为传递函数 $W(s)$ 的分子多项式和分母多项式.使

$$P(s)=a_0+a_1s+a_2s^2+\cdots+a_ns^n=0$$

的点即为传递函数的极点.极点可以是实数,若为复数则以共轭对形式出现.极点在 s 复数平面上的分布,决定了系统的稳定性并影响到其他性能,因此在控制系统的分析和设计中,要特别重视极点位置在复数平面上的配置.

传递函数的零点(zeros of transfer function) 经典控制理论的基本概念.是使系统的传递函数取零的变量(频率变量)值.若线性定常系统的传递函数为

$$W(s)=\frac{b_0+b_1s+b_2s^2+\cdots+b_ms^m}{a_0+a_1s+a_2s^2+\cdots+a_ns^n},$$

使

$$b_0+b_1s+b_2s^2+\cdots+b_ms^m=0$$

的点即为传递函数的零点.零点可以是实数,若为复数则以共轭对形式出现.

梅森增益公式(Mason gain-formula) 计算总增益的公式.指系统信号流图中用以确定输入节点与输出节点之间总增益的一个计算公式.梅森增益公式可表示为

$$P=\frac{1}{\Delta}\sum_k P_k\Delta_k,$$

其中 P 为总增益,Δ 是流图的特征式,可用下式表示:

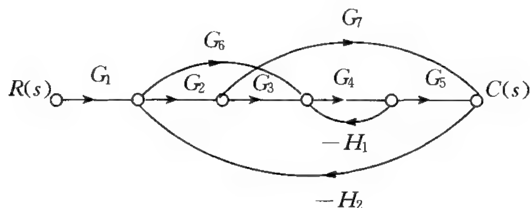
$$\Delta=1-\sum_a L_a+\sum_{b,c}L_bL_c-\sum_{d,e,f}L_dL_eL_f+\cdots,$$

式中 $\sum_a L_a$ 是所有不同回路的增益之和,$\sum_{b,c}L_bL_c$ 是每两个互不接触回路增益乘积之和,$\sum_{d,e,f}L_dL_eL_f$ 是每三个互不接触回路增益乘积之和. P_k 是第 k 条前向通路的增益,Δ_k 是在除去与第 k 条前向通路相接触的回路后的 Δ 值.对于如下图所示的系统,输入量 $R(s)$ 与输出量 $C(s)$ 之间有三条前向通路,这些前向通路的增益分别为:

$$P_1=G_1G_2G_3G_4G_5,$$
$$P_2=G_1G_6G_4G_5,$$
$$P_3=G_1G_2G_7.$$

图中存在四个不同回路,这四个回路的增益分别为:

$$L_1=-G_4H_1,\quad L_2=-G_2G_7H_2,$$
$$L_3=-G_6G_4G_5H_2,\quad L_4=-G_2G_3G_4G_5H_2.$$



因回路 L_1 不接触回路 L_2,故特征式 Δ 为

$$\Delta=1-(L_1+L_2+L_3+L_4)+L_1L_2.$$

从 Δ 中去掉与前向通路 P_1 相接触的回路,可得

$$\Delta_1=1.$$

类似地,可求得

$$\Delta_2=1,\quad \Delta_3=1-L_1.$$

因此可用梅森公式求出输入 $R(s)$ 与输出 $C(s)$ 间的总增益(闭环传递函数)为

$$\frac{C(s)}{R(s)}=\frac{1}{\Delta}(P_1\Delta_1+P_2\Delta_2+P_3\Delta_3)$$
$$=\frac{G_1G_2G_3G_4G_5+G_1G_6G_4G_5+G_1G_2G_7}{(1+G_4H_1)_2}.$$

控制系统的典型环节(typical loops for control systems) 以系统传递函数可能分解的最简形式表征的基本组成单元.尽管自动控制系统中物理过程的形式可千变万化,如机械、液压、传热、电路、电机、电子等,但其变换模型的数学性质和动态性能的数学形式具有很强的相似性.因此,如果把某个物理单元的作用看成是整个系统的一个组成部分,并抽象为一个简单的数学模型,则一个复杂的物理过程就可表示成多个简单模型按不同方式的连结.对于线性定常系统的数学模型,它的基本组成单元的数学模型已标准化,具有有限类型,故称为典型环节.以这些环节的数学形式和在系统中表现的性能加以分

类，有比例环节、微分环节、积分环节、惯性环节、振荡环节等.

比例环节(proportion loop)　控制系统的一类典型环节. 比例环节的传递函数

$$W(s) = k(\text{常数}).$$

比例环节的输入量和输出量具有完全相同的时间函数，只是幅度上有一固定的数值比例，并称 k 为增益. 工程上比例环节常由宽频带放大器组成. 在简单情况下，也可以由电阻分压器组成.

积分环节(integration loop)　控制系统的一类典型环节. 积分环节的传递函数

$$W(s) = \frac{k}{s},$$

其中 s 为拉普拉斯变换中的算子变量，k 为一比例常数. 积分环节的输出量与输入量的时间积分值成比例，积分环节在消除控制系统中的静态误差方面极为有效，常使用在距离、转角等物理量的精确无静差跟踪上. 最简单的积分环节可由一个电容充电过程来表征. 实际使用时，常由宽频带放大器经过电容器反馈而构成.

微分环节(differentiation loop)　控制系统的一类典型环节. 微分环节的传递函数

$$W(s) = ks,$$

其中 s 为拉普拉斯变换中的算子变量，k 为一比例常数. 微分环节的输出量与输入量对时间变量的导数值成比例. 微分作用反映其输入信号的变化速率. 因此，将微分环节引入控制系统中，可使系统的输出及早得到修正. 最简单的微分环节可由信号经电容器隔离后的输出来表征. 工程应用中，则由在宽频带放大器电路中连结电阻、电容反馈支路后组成. 微分环节对于输入中的高频干扰十分敏感，常导致信号被干扰所掩没.

惯性环节(inertia loop)　控制系统的一类典型环节. 惯性环节的传递函数

$$W(s) = \frac{1}{Ts + 1},$$

其中 s 为拉普拉斯变换中的算子变量，T 为时间常数. 惯性环节输出量的变化滞后于其输入量变化. 物理过程的控制中许多过程属于惯性环节，例如，炉子温度的调节、容器水位的调节中都会呈现出系统的惯性. 惯性环节相应的微分方程解式是负指数时间函数，故惯性环节对系统的稳定性不产生影响，仅造成系统反应速度缓慢. 惯性环节又称非周期环节.

振荡环节(oscillation loop)　控制系统的一类典型环节. 振荡环节的传递函数

$$W(s) = \frac{1}{T^2s^2 + 2\zeta Ts + 1},$$

其中 s 为拉普拉斯变换中的算子变量，T 和 ζ 为常量，分别称为时间常数和阻尼系数. 振荡环节的输入量发生变化时，输出量常会呈现周期性变化，其频率仅与环节有关，与信号的幅值和变化速度无关. 振荡环节对应于二阶常微分方程，它具有振荡特征的充分必要条件是 $\zeta<1$，即传递函数的分母多项式有虚根或共轭复数根. 当 $\zeta\leqslant 0$ 时，振荡环节出现持续以至发散的振荡，系统为不稳定. 当 $1>\zeta>0$ 时，振荡环节相应于衰减性振荡，系统可稳定工作，但一般希望对它的振荡频率、幅值和衰减速度进行限制，以免造成不良的后果.

过渡过程(transition process)　对系统动态特征的一种刻画. 它是系统在输入作用下输出量由原来稳态变化到新的稳态的过程. 过渡过程对应于系统数学模型微分方程的自由解，其形态只由系统结构和参数所决定. 过渡过程对系统的性能有着重要影响. 在分析过渡过程时，通常考虑两种最典型的突变输入量.

第一种输入量是单位阶跃函数，即在某一时刻突然从 0 跳变为恒定值 1，相应输出量的过渡过程称为单位阶跃响应.

第二种输入量是单位脉冲函数，即为一个无限窄的脉冲且脉冲在时间轴上的积分值为 1，相应输出量的过渡过程称为脉冲响应.

超调量(overshoot)　反映系统过渡过程性能的一个指标. 系统在过渡过程中输出量最大值对稳态值的相对偏差值. 通常采用两者的偏差对稳态值的百分比来衡量. 超调量反映了系统“超出需要”的调节能力. 超调量过大将对某些设备、元件产生破坏作用，并使衰减过程变长. 不存在超调量的系统会导致输出的变化缓慢，过渡过程的时间增长. 在许多情况下，往往允许一定比例的超调量(5%～15%)存在. 超调量的存在反映出系统的传递函数中包含有振荡环节.

一阶系统(1st-order system)　控制系统按数学模型分类时的一种形式. 是用数学模型可表示为一阶线性常微分方程的系统. 不论系统的具体物理过程为何，数学上一阶系统的解具有最简单的指数函数形式. 其中幂指数可以取正值或负值. 幂指数为正时，函数值将随着时间的增长而无限增大. 这类发散过程一般不符合各种物理系统的正常工作状态，也是控制系统中所不希望的结果. 幂指数为负的一阶系统具有指数型的收敛过程，因此是自动控制系统中最受重视的一类系统. 一阶系统动态过程的收敛速度，取决于幂指数中所包含的时间常数的数值，一般称输出量从起始增加到最终稳定值的 64%的时间间隔为系统的时间常数. 时间常数是表征一阶系统的一个重要参数.

二阶系统(2nd-order system)　控制系统按数

学模型分类时的一种形式. 是用数学模型可表示为二阶线性常微分方程的系统. 二阶系统的解的形式,可由对应传递函数 $W(s)$ 的分母多项式 $P(s)$ 来判别和划分. $P(s)$ 的一般形式为变换算子 s 的二次三项代数式,经标准化后可记为

$$P(s) = 1 + a_1 s + a_2 s^2.$$

代数方程 $P(s)=0$ 的根,可能出现四种情况:

1. 两个实根的情况,对应于两个串联的一阶系统. 如果两个根都是负值,就为非周期性收敛的稳定情况.

2. 当 $a_1=0, a_2>0$,即一对共轭虚根的情况,将引起频率固定的等幅振荡,是系统不稳定的一种表现.

3. 当 $a_1<0, a_1^2-4a_2<0$,即共轭复根有正实部的情况,对应于系统中发生发散型的振荡,也是不稳定的一种表现.

4. 当 $a_1>0, a_1^2-4a_2<0$,即共轭复根有负实部的情况,对应于收敛型振荡,且实部和虚部的数值比例对输出过程有很大的影响. 一般以阻尼系数 ζ 来表征,常取

$$\zeta = \frac{a_1}{(2\sqrt{a_2})}$$

在 0.4~0.8 之间为宜. 当 $\zeta>0.8$ 后,振荡的作用就不显著,输出的速度也比较慢. 而 $\zeta<0.4$ 时,输出量就带有明显的振荡和较大的超调量,衰减也较慢,这也是控制系统中所不希望的.

单输入-单输出最小相位系统(single input-single output minimum phase system) 一类控制系统. 是零点和极点均分布在左半开 s 复平面上的单输入-单输出的线性定常系统. 反之,如果在右半 s 平面上分布有零点或极点,则称其为非最小相位系统. 对于最小相位系统,在具有相同幅频特性的一类系统中,当 ω 从 0 变化至 ∞ 时,系统的相角变化范围最小,"最小相位"之名即源于此. 与之相比,而非最小相位系统的相角变化范围则要大. 考虑如下两个系统:

$$G_1(s) = \frac{1+T_2 s}{1+T_1 s},\quad G_2(s) = \frac{1-T_2 s}{1+T_1 s},$$

其中 $T_1>T_2>0$. $G_1(s)$ 为最小相位系统,$G_2(s)$ 为非最小相位系统. 两系统的幅频特性均为

$$\sqrt{\frac{1+\omega^2 T_2^2}{1+\omega^2 T_1^2}},$$

相频特性分别为

$$\varphi_1(\omega) = \arctan \omega T_2 - \arctan \omega T_1,$$

$$\varphi_2(\omega) = \arctan(-\omega T_2) - \arctan \omega T_1.$$

在 ω 从 0 至 ∞ 范围内,$G_1(s)$ 的相角变化很小,而 $G_2(s)$ 的相角则从 0°变化到 −180°. 又如,根据定义,时滞环节是非最小相位的,相角将随 ω 增大而增大.

单输入-单输出非最小相位系统(single input-single output non-minimum phase system) 见"单输入-单输出最小相位系统".

控制系统的稳定性(stability of control system) 控制系统的一种特性. 它反映受控系统运动过程相对于平衡状态的渐近收敛性或有界性. 稳定性是保证系统能正常工作的最基本的条件,因而是对控制系统所要求的最重要的性质. 常用的控制系统稳定性概念有两种:一种是闭环系统的李亚普诺夫稳定性,它研究系统的初始状态受到扰动后,系统的运动是否收敛(参见"李亚普诺夫稳定性");另一种是有界输入-有界输出稳定性(简称输入-输出稳定性),它研究系统受到任一有界输入后,其输出是否也有界(参见"输入-输出稳定性"). 对于能控、能观的线性定常系统(参见"能控性"和"能观测性"),这两种稳定性是等价的.

通常用李亚普诺夫第二方法研究控制系统的李亚普诺夫稳定性;用小增益定理和无源性定理研究控制系统的输入-输出稳定性. 对于连续线性定常系统,可采用劳斯判据、赫尔维茨判据、奈奎斯特判据等时间域或频率域判据来判定系统的稳定性. 对于离散线性定常系统,可采用舒尔判据判定系统的稳定性.

劳斯判据(Routh criterion) 关于多项式方程具有正实部根的一个判据. 它根据多项式系数经简单代数运算来判别多项式方程是否包含正实部根. 考虑以下 s 的多项式方程

$$a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n = 0,$$

其根的实部全为负的必要条件之一是系数 $a_0, a_1, \cdots, a_n$ 的符号相同. 如果所有系数都是正值,构造如下的劳斯阵列

$$\begin{array}{ccccc}
a_0 & a_2 & a_4 & a_6 & \cdots \\
a_1 & a_3 & a_5 & a_7 & \cdots \\
b_1 & b_2 & b_3 & b_4 & \cdots \\
c_1 & c_2 & c_3 & c_4 & \cdots \\
d_1 & d_2 & d_3 & d_4 & \cdots \\
e_1 & e_2 & e_3 & e_4 & \cdots \\
f_1 & f_2 & f_3 & f_4 & \cdots \\
g_1 & g_2 & g_3 & g_4 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots
\end{array}$$

其中系数 $b_1, b_2, b_3, \cdots$ 按下列公式逐行计算:

$$b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1},$$

$$b_2 = \frac{a_1 a_4 - a_0 a_5}{a_1},$$

$$b_3=\frac{a_1a_6-a_0a_7}{a_1},$$

………………

$$c_1=\frac{b_1a_3-a_1b_2}{b_1},$$

$$c_2=\frac{b_1a_5-a_1b_3}{b_1},$$

$$c_3=\frac{b_1a_7-a_1b_4}{b_1},$$

………………

$$d_1=\frac{c_1b_2-b_1c_2}{c_1},$$

$$d_2=\frac{c_1b_3-b_1c_3}{c_1},$$

………………

劳斯判据指出：多项式方程实部为正的根的个数，等于劳斯阵列中第一列系数符号改变的次数. 由劳斯判据推知，多项式方程的全部根都位于左半开 s 平面的充分必要条件是：方程全部系数都是正值，且劳斯阵列第一列中所有项都为正值. 劳斯判据是劳斯(Routh, E. J.)于 1877 年提出的.

赫尔维茨判据(Hurwitz criterion)　关于多项式方程具有正实部根的一个判据. 它根据多项式系数经代数运算来判定多项式方程是否有不稳定根. 赫尔维茨判据指出，对 s 的多项式方程

$$a_0s^n+a_1s^{n-1}+\cdots+a_{n-1}s+a_n=0,$$

其所有根的实部均为负的充分必要条件为：

1. 所有系数 $a_0,a_1,\cdots,a_n$ 均为正.
2. 如下 n 个赫尔维茨行列式均为正：

$$H_i=\begin{vmatrix} a_1 & a_3 & a_5 & \cdots & a_{2i-1} \\ a_0 & a_2 & a_4 & \cdots & a_{2i-2} \\ 0 & a_1 & a_3 & \cdots & a_{2i-3} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_i \end{vmatrix}\quad (i=1,2,\cdots,n).$$

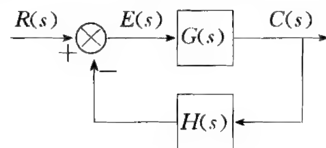
赫尔维茨判据是赫尔维茨(Hurwitz, A.)于 1895 年提出的.

***D* 域划分**(*D*-division)　一种刻画多项式方程根分布的几何方法. 是把多项式方程根平面的虚轴映射到方程系数空间中一定区域的过程. 通常，把这一区域称为 D 域. 考虑如下的多项式方程

$$a_0s^n+5s^{n-1}+\cdots+10s+a_n=0,$$

其系数除 a_0 和 a_n 外均为已知定值. 设多项式方程在左半闭 s 平面有 k 个根，在右半闭 s 平面有 $n-k$ 个根. 系数 a_0 和 a_n 在一定范围内变化，不会导致根平面上虚轴左右两边根的数目的变化. 因此，可在 a_0 和 a_n 平面上划分一个区域，域中每一点所对应的多项式方程都在左半 s 平面有 k 个根，在右半 s 平面有 $n-k$ 个根，称之为 $D(k)$ 域. 多项式中的未定系数可以是任意两个，也可以是大于两个，对后一情况 $D(k)$ 域是在多维空间中由超曲面作为分界面. 系数空间中穿越 $D(k)$ 域边界，对应于根平面中有根穿越虚轴. 因此，只要以 $j\omega$ 代替 s，并令 ω 由 $-\infty$ 变到 $+\infty$，则所考虑的多项式系数的变化轨迹就是 $D(k)$ 域的边界. 工程问题中，所关心的是 $D(n)$ 域，即系统特征多项式方程的所有根都在左半开 s 平面时对应的系数空间的 D 域. 当所关心的系数在 D 域内变化时，系统的稳定性保持不变. 此外，也可相对于其他任何参量空间进行 D 域划分，前提是已知多项式系数是这些参量的函数.

稳态误差系数(steady state error coefficient)　控制系统的稳态精度的一种度量. 对于不同类型的输入，一个控制系统的稳态误差一般是不同的；对于某一类型的输入，一控制系统是否产生稳态误差，只取决于系统的开环传递函数形式. 考虑图中所示控制系统的开环传递函数



控制系统方框图

$$G(s)H(s)=\frac{K(T_as+1)(T_bs+1)\cdots(T_ms+1)}{s^N(T_1s+1)(T_2s+1)\cdots(T_ps+1)},$$

分别称 $N=0,N=1,N=2,\cdots$ 时的系统为 0 型、1 型、2 型、…系统. 系统的误差信号 $e(t)$ 和输入信号 $r(t)$ 之间的传递函数为

$$\frac{E(s)}{R(s)}=\frac{1}{1+G(s)H(s)}.$$

由拉普拉斯变换的终值定理，稳定系统的稳态误差可表示为

$$e_{ss}=\lim_{t\to\infty}e(t)=\lim_{s\to 0}\frac{sR(s)}{1+G(s)H(s)}.$$

定义静态位置误差系数 K_p 为

$$K_p=\lim_{s\to 0}G(s)H(s).$$

对于单位阶跃输入的稳态误差 e_{ss} 可表示为

$$e_{ss}=\frac{1}{1+K_p}=\frac{1}{1+K}\ (0\text{ 型系统}),$$

$$e_{ss}=\frac{1}{1+K_p}=0\ (1\text{ 型或高于 }1\text{ 型的系统}).$$

定义静态速度误差系数 K_v 为

$$K_v=\lim_{s\to 0}sG(s)H(s).$$

对于单位斜坡输入的稳态误差 e_{ss} 可表示为

$$e_{ss}=\frac{1}{K_v}=\infty\quad (0\text{ 型系统}),$$

$$e_{ss}=\frac{1}{K_v}=\frac{1}{K}\quad (1\text{ 型系统}),$$

$$e_{ss}=\frac{1}{K_v}=0\quad (2\text{ 型或高于 }2\text{ 型的系统}).$$

定义静态加速度误差系数 K_a 为

$$K_a = \lim_{s \to 0} s^2 G(s) H(s).$$

对于单位加速度输入的稳态误差 e_{ss} 可表示为

$$e_{ss} = \frac{1}{K_a} = \infty \quad (0\ 型和\ 1\ 型系统),$$

$$e_{ss} = \frac{1}{K_a} = \frac{1}{K} \quad (2\ 型系统),$$

$$e_{ss} = \frac{1}{K_a} = 0 \quad (3\ 型或高于\ 3\ 型的系统).$$

静态误差系数 K_p, K_v, K_a 可用来表征系统对减小或消除稳态误差的能力. 为了改善稳态性能,可以在控制系统的前向通路中增加一个或几个积分环节,使系统的型号增加,但这会带来稳定性问题,设计时需要兼顾考虑.

无静差系统(systems without steady state error) 控制系统按稳态误差分类时的一种形式. 是对于某类输入的稳态误差为零的系统. 1 型或高于 1 型的系统对于单位阶跃输入为无静差系统;2 型或高于 2 型的系统对于单位阶跃输入、斜坡输入都是无静差系统;3 型或高于 3 型的系统对于单位阶跃输入、斜坡输入、加速度输入都是无静差系统.

控制工程上,常称 1 型系统为 1 阶无静差系统,2 型系统为 2 阶无静差系统,3 型系统为 3 阶无静差系统,意即它们分别对单位阶跃、单位斜坡及单位加速度输入是无静差系统.

有静差系统(systems with steady state error) 控制系统按稳态误差分类时的一种形式. 对于某类输入,稳态误差不为零的系统称为有静差系统. 对于单位阶跃输入,0 型系统是有静差系统;对于单位斜坡输入,0 型和 1 型系统都是有静差系统;对于单位加速度输入,0 型、1 型和 2 型系统都是有静差系统.

动态误差系数(dynamic error coefficient) 控制系统稳态精度的一种度量. 静态误差系数只适于刻画稳态误差是常数的系统,动态误差系数可用来描述系统稳态误差随时间的变化规律. 研究"稳态误差系数"条目图中所示的闭环系统,其误差信号 $e(t)$ 与输入信号 $r(t)$ 间的传递函数为

$$\frac{E(s)}{R(s)} = \frac{1}{1 + G(s)H(s)}.$$

把 $E(s)/R(s)$ 展开为 s 的升幂级数:

$$\frac{E(s)}{R(s)} = \frac{1}{k_1} + \frac{1}{k_2}s + \frac{1}{k_3}s^2 + \cdots,$$

此级数的系数 $k_1, k_2, k_3, \cdots$ 被定义为系统的各阶动态误差系数,即 k_1 为动态位置误差系数,k_2 为动态速度误差系数,k_3 为动态加速度误差系数等. 考虑开环传递函数 $G(s)H(s)$ 有如下形式的 N 型系统:

$$G(s)H(s) = \frac{K(T_a s+1)(T_b s+1)\cdots(T_m s+1)}{s^N(T_1 s+1)(T_2 s+1)\cdots(T_p s+1)}.$$

闭环系统的动态误差系数由下列各式给出:

$$k_n = \infty \qquad (\forall n < N),$$

$$k_n = \lim_{s \to 0} s^N G(s) \quad (n = N),$$

$k_n(\forall n > N)$ 可由 $E(s)/R(s)$ 在原点附近的展开式确定. 按 $E(s)/R(s)$ 在原点附近的展开式,$E(s)$ 可改写为

$$E(s) = \frac{1}{k_1}R(s) + \frac{1}{k_2}sR(s) + \frac{1}{k_3}s^2R(s) + \cdots,$$

此级数的收敛域是 $s=0$ 的邻域,这相当于时域内的 $t \to \infty$. 假定系统的所有初始条件为零,且设 $r(t)$ 的各阶导数连续,则对应的稳态误差函数可表达为

$$\lim_{t \to \infty} e(t) = \frac{1}{k_1}r(t) + \frac{1}{k_2}\dot{r}(t) + \frac{1}{k_3}\ddot{r}(t) + \cdots,$$

这样,由输入函数及其各阶导数所引起的稳态误差可根据动态误差系数求出. 动态误差系数为计算任意输入作用下系统稳态误差时间函数提供了一种简单方法,可避免直接求解系统的微分方程. 动态误差系数中动态的含义不是指系统过渡过程中的瞬态误差,而是指系统稳态误差随时间的变化.

误差积分准则(error integral criterion) 控制系统的一类评价准则或设计准则. 误差积分准则既可对一给定系统进行性能评价,又可作为控制器的设计准则. 误差积分准则可分为以下四种:

1. 平方误差积分准则,

$$J_1 = \int_0^\infty e^2(t)\mathrm{d}t.$$

2. 时间乘平方误差积分准则,

$$J_2 = \int_0^\infty te^2(t)\mathrm{d}t.$$

3. 绝对误差积分准则,

$$J_3 = \int_0^\infty |e(t)|\mathrm{d}t.$$

4. 时间乘绝对误差积分准则,

$$J_4 = \int_0^\infty t|e(t)|\mathrm{d}t.$$

上述各误差积分准则中,$e(t)$ 定义为

$$e(t) = y_r(t) - y(t)$$

或

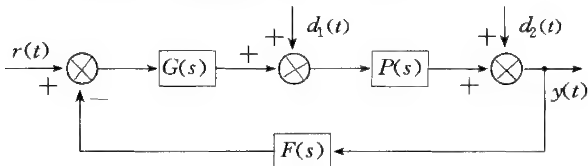
$$e(t) = y(\infty) - y(t),$$

其中 $y_r(t)$ 为设计者期望的输出,$y(t)$ 是系统实际输出,$y(\infty)$ 是系统输出的稳态值. 各种误差积分准则中的积分上限也可用足够大的 T 来代替.

平方误差积分准则着重权衡大的误差,而较少考虑小的误差. 用这种准则设计的控制系统可使系统的大起始误差迅速减小,系统的响应迅速且呈现出振荡性,但系统的相对稳定性较差. 平方误差积分准则等同于工程上的能量消耗最小准则,具有实际意义,但选择性很差,即对于某些性能差别较大的两个系统,其平方误差积分值可能相差很小. 时间乘平

方误差积分准则对系统响应中大的起始误差考虑较少，而着重权衡瞬态响应后期出现的误差. 与平方误差积分准则相比，这个准则具有较好的选择性. 绝对误差积分准则是最易应用的一种性能指标，基于这种准则所设计的系统具有适当的阻尼和令人满意的瞬态响应. 这种准则的选择性不太好，且对于过欠阻尼和过阻尼系统，应用这种准则很难获得满意的设计结果. 这种设计准则应用于宇宙飞船时，即燃料消耗最小准则，具有实际意义. 时间乘绝对误差积分准则对系统响应中大的起始误差考虑较少，而着重权衡瞬态响应后期出现的误差. 应用这种准则设计的系统，其超调量少，振荡有足够阻尼. 此准则还具有良好的选择性. 误差积分准则是二次型最优控制问题的最早雏形.

灵敏度(sensitivity)　系统特性对系统参数或外扰动变化反应的灵敏程度. 研究下图所示控制系统. $P(s)$为被控对象，$G(s)$为前馈控制环节，$F(s)$为反馈控制环节，$d_1(t)$和$d_2(t)$是不同位置的外扰，$r(t)$和$y(t)$是系统的输入和输出信号.



控制系统结构图

系统的总传递函数

$$T(s)=\frac{G(s)P(s)}{1+F(s)G(s)P(s)}.$$

假设$P(s)$变化$\delta P(s)$，把$T(s)$展开为泰勒级数：

$$T(s,P+\delta P)=T(s,P)+\frac{\partial T(s)}{\partial P(s)}\delta P+\cdots,$$

由于δP为小量，略去高阶项，可得

$$\delta T(s)=\frac{\partial T(s)}{\partial P(s)}\delta P(s).$$

系统的灵敏度函数$S_P^T(s)$相应地定义为

$$S_P^T(s)=\frac{\dfrac{\delta T(s)}{T(s)}}{\dfrac{\delta P(s)}{P(s)}}=\frac{1}{1+F(s)G(s)P(s)}.$$

由灵敏度函数，可由对象的微小相对变化，得到系统总传递函数的相对变化：

$$\frac{\delta T(s)}{T(s)}=S_P^T(s)\,\frac{\delta P(s)}{P(s)}.$$

令$s=\mathrm{j}\omega$，则有

$$S_P^T(\mathrm{j}\omega)=\frac{1}{1+F(\mathrm{j}\omega)G(\mathrm{j}\omega)P(\mathrm{j}\omega)}.$$

画出$S_P^T(\mathrm{j}\omega)$的对数频率特性曲线，可以直观地了解系统在不同频段对对象变化的灵敏度大小. 若$P(s)$是参数p_i的函数，$G(s)$是参数g_i的函数，$F(s)$是参数f_i的函数，则可得系统总传递函数对参数p_i，g_i和f_i的灵敏度函数

$$S_{p_i}^T(s)=S_P^T(s)S_{p_i}^P(s)=S_P^T(s)\frac{\dfrac{\delta P(s)}{P(s)}}{\dfrac{\delta p_i}{p_i}},$$

$$S_{g_i}^T(s)=S_G^T(s)S_{g_i}^G(s)=S_G^T(s)\frac{\dfrac{\delta G(s)}{G(s)}}{\dfrac{\delta g_i}{g_i}},$$

$$S_{f_i}^T(s)=S_F^T(s)S_{f_i}^F(s)=S_F^T(s)\frac{\dfrac{\delta F(s)}{F(s)}}{\dfrac{\delta f_i}{f_i}}.$$

令$s=\mathrm{j}\omega$，可画出$S_{p_i}^T(\mathrm{j}\omega)$，$S_{g_i}^T(\mathrm{j}\omega)$，$S_{f_i}^T(\mathrm{j}\omega)$的对数频率特性曲线，能直观地了解系统在不同频段的参数变化的灵敏度大小. 由图可得

$$Y(s)+\delta Y(s)=\frac{G(s)P(s)}{1+F(s)G(s)P(s)}R(s)+\frac{D_1(s)}{1+F(s)G(s)P(s)}.$$

从而外扰$D_1(s)$所引起的输出增量为

$$\delta Y(s)=\frac{D_1(s)}{1+F(s)G(s)P(s)}=S_P^T(s)D_1(s),$$

表明灵敏度函数$S_P^T(s)$也是系统输出对外扰$D_1(s)$的灵敏度函数. 由图还可得

$$Y(s)+\delta Y(s)=\frac{G(s)P(s)}{1+F(s)G(s)P(s)}R(s)+\frac{P(s)}{1+F(s)G(s)P(s)}D_2(s).$$

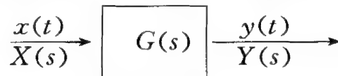
从而

$$\delta Y(s)=\frac{P(s)}{1+F(s)G(s)P(s)}D_2(s)=S_P^T(s)P(s)D_2(s).$$

表明系统输出对外扰$D_2(s)$的灵敏度函数为$S_P^T(s)P(s)$.

灵敏度函数(sensitivity function)　见“灵敏度”.

频率响应(frequency response)　也称频率特性. 控制系统的一种频域特性. 是线性定常系统在不同频率谐波输入下稳态响应的一种特征. 频率响应是以频率ω为自变量的复变函数，通常也称为频率特性. 系统的频率响应既可由系统的传递函数计算得到，也可通过实验方法获得. 利用频率响应，可以对系统的性能进行分析(包括稳定性、瞬态响应等)，也可对系统进行设计，使控制系统达到理想的性能. 考虑下图所示线性定常系统，其中输入为



线性定常系统

$x(t)$，输出为$y(t)$，传递函数为$G(s)$. 对谐波输入而

言，如图所示的系统输出、输入信号的幅值与相位之间有如下关系：

$$|G(\mathrm{j}\omega)| = \left|\frac{Y(\mathrm{j}\omega)}{X(\mathrm{j}\omega)}\right|,$$

$$\arg G(\mathrm{j}\omega) = \arg \frac{Y(\mathrm{j}\omega)}{X(\mathrm{j}\omega)}.$$

系统的频率响应由 $G(\mathrm{j}\omega)$ 得到

$$\frac{Y(\mathrm{j}\omega)}{X(\mathrm{j}\omega)} = G(\mathrm{j}\omega),$$

即在系统的传递函数 $G(s)$ 中以 $\mathrm{j}\omega$ 代替 s 来求得. 频率响应采用它的幅值和相角来表示，它们是以频率 ω 为自变量的实变函数. 因 $G(s)$ 是 s 的实系数有理函数，故 $G(\mathrm{j}\omega)$ 的图形在 $(-\infty,0)$ 和 $(0,+\infty)$ 是关于实轴对称的，一般只需讨论 ω 由 0 变至 $+\infty$ 的情形. 频率响应采用如下三种表达形式：

1. 奈奎斯特图.
2. 伯德图.
3. 尼柯尔斯图.

频率响应法是针对线性定常系统的一种分析和设计方法. 对于二阶系统，系统的频率响应与过渡过程性能指标之间具有确定的对应关系；对于高阶系统，两者之间存在近似关系. 当系统在某些频段内存在严重噪声干扰时，应用频率响应法可设计出有效抑制噪声干扰的控制系统.

奈奎斯特图(Nyquist plot)　系统频率响应的一种图示方法. 即当 ω 由零变化到无穷大时，奈奎斯特图表示为复平面上不同频率 ω 时 $G(\mathrm{j}\omega)$ 幅值与 $G(\mathrm{j}\omega)$ 相角的关系图，即以 ω 为参变量的矢量 $G(\mathrm{j}\omega)$ 的轨迹. 奈奎斯特图又称极椭圆.

奈奎斯特图的优点是能直观地显示出不同频率下 $G(\mathrm{j}\omega)$ 的幅值与相角. 为绘制奈奎斯特图，需要计算出每个频率值 ω 所对应的 $G(\mathrm{j}\omega)$ 的幅值 $|G(\mathrm{j}\omega)|$ 和相角 $\arg G(\mathrm{j}\omega)$ 或 $G(\mathrm{j}\omega)$ 的实部与虚部. 图 1 所示为几种简单传递函数的奈奎斯特图.

研究系统传递函数形如

$$G(\mathrm{j}\omega) = \frac{K(T_a\mathrm{j}\omega + 1)(T_b\mathrm{j}\omega + 1)\cdots}{(\mathrm{j}\omega)^N(T_1\mathrm{j}\omega + 1)(T_2\mathrm{j}\omega + 1)\cdots} = \frac{b_0(\mathrm{j}\omega)^m + b_1(\mathrm{j}\omega)^{m-1} + \cdots}{a_0(\mathrm{j}\omega)^n + a_1(\mathrm{j}\omega)^{n-1} + \cdots}$$

的频率响应，式中 $n>m$.

$G(\mathrm{j}\omega)$ 的奈奎斯特图在形状上有如下特点：

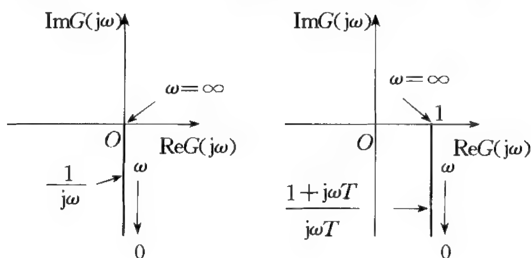


图 1　$G(\mathrm{j}\omega)$ 在复平面上的轨迹

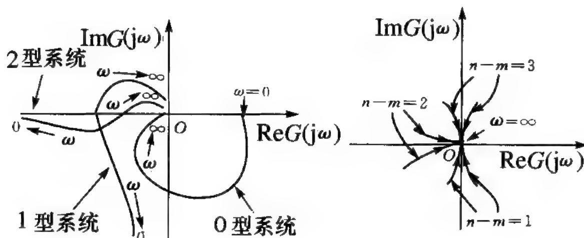


图 2　　　　图 3

1. $N=0$，即 0 型系统. 奈奎斯特图的起始点(对应于 $\omega=0$)是正实轴上的有限值，曲线在 $\omega=0$ 处的切线垂直于实轴，曲线的终点(对应于 $\omega=\infty$)位于原点，且曲线与实轴相切.

2. $N=1$，即 1 型系统. 当 $0\leqslant\omega<\infty$ 时，$G(\mathrm{j}\omega)$ 相角中的 $-90°$ 是由分母中的 $\mathrm{j}\omega$ 项产生的. 在低频段，曲线渐近平行于负虚轴. 在 $\omega=\infty$ 处，曲线收敛于原点并与虚轴相切.

3. $N=2$，即 2 型系统. 当 $0\leqslant\omega<\infty$ 时，$G(\mathrm{j}\omega)$ 相角中的 $-180°$ 是由分母中的 $(\mathrm{j}\omega)^2$ 项产生的. 在 $\omega=0$ 处，$G(\mathrm{j}\omega)$ 的幅值为无穷大，相角为 $-180°$，在低频处，曲线渐近平行于负实轴，在 $\omega=\infty$ 处，曲线收敛于原点并与实轴相切.

图 2 给出了 0 型、1 型和 2 型低频段曲线的一般形状. 图 3 给出了 $\omega\to\infty$ 时曲线的收敛情况.

伯德图(Bode plot)　系统频率响应的一种图

示方法．伯德图由幅值图和相角图组成，两者都按频率的对数分度绘制．伯德图常也称为对数坐标图．

对数幅值的标准表达式为 $20\lg|G(j\omega)|$，单位是分贝，缩写为 dB．相角的单位是度．通常，伯德图被画于半对数坐标纸上，横坐标频率采用对数分度，纵坐标幅值或相角采用线性分度．

伯德图的主要优点一是可将幅值相乘转化为对数幅值相加；二是这种方法建立在渐近近似的基础上，可采用简便方法绘制近似的对数幅值曲线进行

典型环节	(1)比例环节 K (2)延迟环节 $e^{\tau s}$	(1)积分环节 $1/s$ (2)微分环节 s
伯德图	幅值/dB：K (1)；0，1，10，ω；(2) 相角：1 (1)；0，10，ω；(2)	幅值/dB：20，(1)，(2)；0，1，10，ω；−20 相角：(2)；90°；0°，1，10，ω；−90°，(1)
典型环节	(1)惯性环节 $1/(Ts+1)$ (2)一阶微分环节 $Ts+1$ (3)不稳定惯性环节 $1/(Ts-1)$ (4)不稳定一阶微分环节 $Ts-1$	(1)振荡环节 $1/[(s/\omega_n)+2\zeta(s/\omega_n)+1]$ (2)二阶微分环节 $(s/\omega_n)^2+2\zeta(s/\omega_n)+1$ (3)不稳定振荡环节 $1/[(s/\omega_n)^2-2\zeta(s/\omega_n)+1]$ (4)不稳定二阶微分环节 $(s/\omega_n)^2-2\zeta(s/\omega_n)+1$
伯德图	幅值/dB：20，(2) (4)；1，10，ωT；−20，(1) (3) 相角：90°，(2)，(4)；1，10，ωT；−90°，(1)，(3)	幅值/dB：40，(2) (4)；1，10，ω/ω_n；−40，(1) (3) 相角：180°，(2)，(4)；ω/ω_n；−180°，(1)，(3)

工程分析和设计；三是低频特性在实际系统中最为重要，对频率采用对数分度可扩展低频段，更便于工程应用．对于前向回路为 $G(s)$ 和反馈回路为 $H(s)$ 的闭环控制系统，将开环传递函数 $G(s)H(s)$ 的分子、分母多项式分别分解成因式，可得到下列七种典型环节：

1．比例环节 K．

2．惯性环节 $\dfrac{1}{Ts+1}$ $(T>0)$．

3．一阶微分环节 $Ts+1$ $(T>0)$．

4．积分环节 $\dfrac{1}{s}$．

5．微分环节 s．

6．振荡环节 $\dfrac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$ $(\omega_n>0,\ 0<\zeta<1)$．

7．二阶微分环节 $\left(\dfrac{s}{\omega_n}\right)^2+2\zeta\dfrac{s}{\omega_n}+1$ $(\omega_n>0,\ 0<\zeta<1)$．

8．延迟环节 $e^{\tau s}$．

典型环节的伯德图如表所示．绘制伯德图的一般步骤为：首先将开环频率特性 $G(j\omega)H(j\omega)$ 改写为基本环节的乘积，画出各基本环节的伯德图，然后把各基本环节伯德图的对数幅值相加，相角相加，就得到 $G(j\omega)H(j\omega)$ 的伯德图．

伯德图是伯德(Bode，H. W.)于 1940 年提出的．

尼柯尔斯图(Nichols chart)　系统频率响应的一种图示方法．尼柯尔斯图是系统频率响应 $G(j\omega)$ 在所需频率范围内由以分贝数表示的对数幅值与相角的一种关系图，纵轴为对数幅值，横轴为相角，频率 ω 作为参变量．尼柯尔斯图也常称为对数幅-相图．利用尼柯尔斯图，可以较快地确定闭环系统的稳定余量，较容易地解决系统校正问题，并由开环频率响应获得闭环频率响应．尼柯尔斯图可直接根据伯德图上的对数幅值和相角的读数来绘制．改变开环频率响应的增益，对应于使尼柯尔斯图曲线向上或向下移动，曲线形状保持不变．尼柯尔斯图上，开环频率响应 $G(j\omega)$ 的曲线与 0 分贝线的交点所对应的相角与 $-180°$ 之差即为闭环系统的相位裕量，与 $-180°$ 线的交点所对应的分贝数即为闭环系统的增益裕量．确定闭环频率响应，需要利用事先作好的通用的幅-相曲线族图，将开环对数幅-相曲线画在其上，由两者的交点即可得到闭环频率响应．这种通用的幅-相曲线族图称为尼柯尔斯图线．

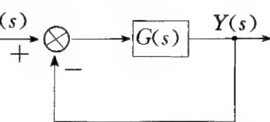


图 1　单位反馈系统

为了导出尼柯尔斯图线的基本关系，考虑图 1 所示的单位反馈系统，其闭环传递函数为

$$T(s)=\frac{G(s)}{1+G(s)},$$

其闭环频率响应可表示为

$$T(j\omega)=\frac{G(j\omega)}{1+G(j\omega)}=M(\omega)e^{j\alpha(\omega)}.$$

其中，$M(\omega)$ 表示闭环频率响应的幅值，$\alpha(\omega)$ 为相角，它们都是频率 ω 的函数．把开环频率响应表示为由幅值和相角表示的复数向量

$$G(j\omega)=|G(j\omega)|e^{j\theta}.$$

将此代入闭环频率响应 $T(j\omega)$ 的表达式：

$$M(\omega)e^{j\alpha(\omega)}=\frac{|G(j\omega)|e^{j\theta}}{1+|G(j\omega)|e^{j\theta}}$$

$$=\left[\frac{e^{-j\theta}}{|G(j\omega)|}+1\right]^{-1}$$
$$=\left[\frac{\cos\theta}{|G(j\omega)|}-\frac{j\sin\theta}{|G(j\omega)|}+1\right]^{-1}.$$

可得 $M(\omega)$ 和 $\alpha(\omega)$ 与 $|G(j\omega)|$ 和 θ 的解析关系：

$$M(\omega)=\left\{\left[1+\frac{1}{|G(j\omega)|^2}+\frac{2\cos\theta}{|G(j\omega)|}\right]^{\frac{1}{2}}\right\}^{-1},$$

$$\alpha(\omega)=-\arctan\frac{\sin\theta}{\cos\theta+|G(j\omega)|}.$$

基于此，以 M 和 α 为参量，构成 G(dB)对 θ 的曲线族图，如图 2 所示，曲线族对称于 $-180°$ 直线.

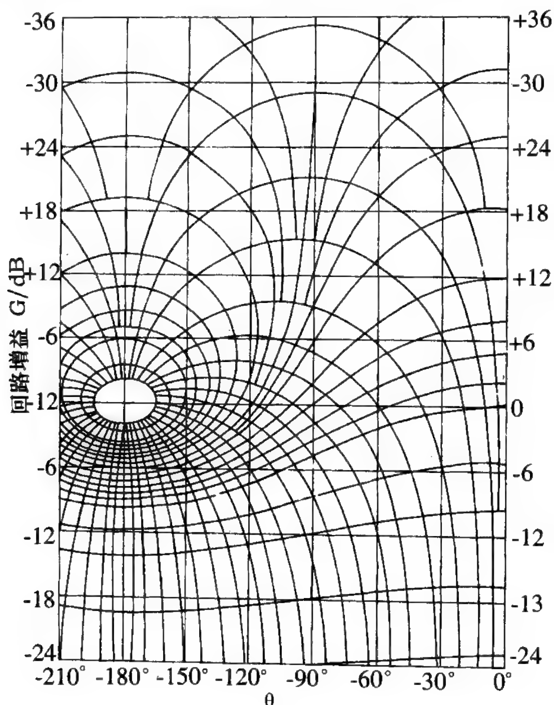


图 2

尼柯尔斯图是尼柯尔斯(Nichols, N. B.)于 1947 年提出的.

增益裕量(gain margin) 系统稳定余量的一种表达形式. 增益裕量定义为系统频率响应 $G(j\omega)$ 的相位等于 $-180°$ 的频率上幅值 $|G(j\omega)|$ 的倒数. 一般地，对"尼柯尔斯图"中所示的单位反馈系统，$G(j\omega)$ 的轨迹越接近于 $-1+j0$ 点，闭环系统响应的振荡性越大，稳定性越差. $G(j\omega)$ 轨迹对 $-1+j0$ 的靠近程度可以用来度量闭环系统的稳定裕量. 稳定裕量以增益裕量和相位裕量的形式来表示. 为分析系统稳定程度，需同时给出这两个量. 定义开环频率响应 $G(j\omega)$ 的相角等于 $-180°$ 时的频率 ω_1 为相位交界频率，增益裕量

$$K_g=\frac{1}{|G(j\omega_1)|}.$$

采用分贝(dB)表示时有

$$K_g(\text{dB})=-20\lg|G(j\omega_1)|.$$

对以分贝表示的增益裕量，如果 $K_g>1$，则增益裕量为正值；如果 $K_g<1$，则增益裕量为负值. 正增益裕量表明系统是稳定的，负增益裕量表明系统是不稳定的. 对于稳定的最小相位系统，增益裕量指出了闭环系统趋于不稳定系统之前，增益所允许的增加值. 对于不稳定系统，增益裕量指出了为使系统稳定，增益所必须的减小值.

下面的图 1，图 2，图 3 中，给出了稳定系统和不稳定系统分别在奈奎斯特图、伯德图和尼柯尔斯图中的增益裕量.

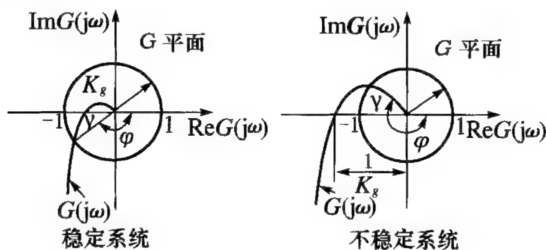


图 1

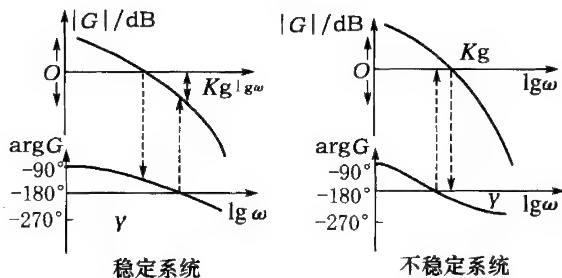


图 2

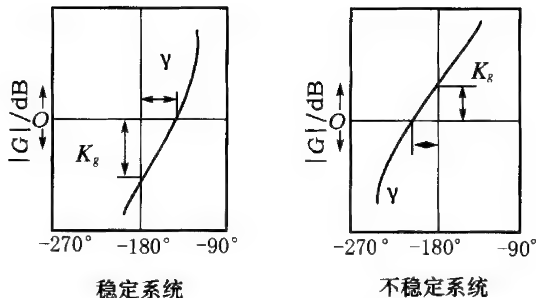


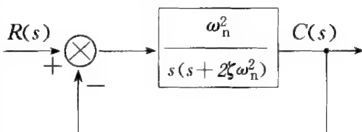
图 3

相位裕量(phase margin) 系统稳定余量的一种表达形式. 相位裕量定义为增益交界频率上，使系统达到不稳定边缘所需要的附加相位滞后量. 所谓增益交界频率是指开环频率响应的幅值 $|G(j\omega)|=1$ 时的频率. 相位裕量 γ 等于 180°加上开环频率响应在增益交界频率上的相角 φ，即 $\gamma=180°+\varphi$.

相位裕量可视 $G(j\omega)$ 取正或取负. 为使最小相位系统稳定，相位裕量必须为正值. 在"增益裕量"条目中的图 1、图 2、图 3 分别给出了稳定系统和不稳定系统在奈奎斯特图、伯德图、尼柯尔斯图上的相位

裕量.控制系统的相位裕量和增益裕量是系统开环频率响应曲线对$-1+j0$点靠近的程度,不仅可用来分析系统的稳定性,也可以作为控制系统的设计准则.单独采用增益裕量或相位裕量,都不足以说明系统的相对稳定性,为了确定系统的稳定性必须同时定出这两个量.对于开环最小相位系统,只有当相位裕量和增益裕量都是正值时,闭环系统才是稳定的.

谐振频率(resonant frequency) 控制系统的一种频率特性.系统闭环频率响应的幅值的最大值所对应的频率.研究如图所示的单位反馈二阶系统.



单位反馈二阶系统

该系统的闭环传递函数为

$$T(s)=\frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2},$$

式中ζ和ω_n分别为阻尼比和无阻尼自然频率.系统的闭环频率响应表示为

$$T(j\omega)=\frac{\omega_n^2}{(j\omega)^2+2\zeta\omega_n(j\omega)+\omega_n^2}=Me^{j\alpha},$$

式中

$$M=\frac{1}{\sqrt{\left(1-\frac{\omega^2}{\omega_n^2}\right)^2+\left(2\zeta\frac{\omega}{\omega_n}\right)^2}},$$

$$\alpha=-\arctan\frac{2\zeta\frac{\omega}{\omega_n}}{1-\frac{\omega^2}{\omega_n^2}}.$$

若M在某一频率处有最大值,则此频率称为谐振频率.可以算出:当$0\leqslant\zeta\leqslant0.707$时,谐振频率$\omega_r$和$M$的最大值$M_r$分别为:

$$\omega_r=\omega_n\sqrt{1-2\zeta^2},\ M_r=\frac{1}{2\zeta\sqrt{1-\zeta^2}}.$$

由于仅当$\zeta\leqslant0.707$时,ω_r才等于实数,所以$\zeta>0.707$时,系统不存在谐振现象.

对于二阶系统,从系统的闭环频率响应的M_r和ω_r,可精确地计算出二阶系统的时间响应.对于高阶系统,若存在一对共轭闭环主导极点,则可将二阶系统的时间响应与频率响应的关系扩展到高阶系统,这时谐振频率ω_r和谐振幅值M_r由对应的高阶系统的共轭闭环主导极点来定出.这时,高阶系统的阶跃瞬态响应与频率响应之间有如下关系:

1. M_r表征系统的相对稳定度.如果M_r的值在1.0~1.4(即0~3dB)范围内,则相当于等效阻尼比ζ为0.4~0.7的范围内,可以获得满意的瞬态性能.当M_r的值大于1.5时,阶跃瞬态响应将出现几次超调振荡.一般地,M_r的值越大,相应的瞬态响应的超调量就越大.

2. ω_r表征瞬态响应的速度.ω_r的值越大,时间响应就越快,即上升时间随ω_r成反比变化.

3. 对于弱阻尼系统,谐振频率ω_r与阶跃瞬态响应的阻尼自然频率ω_a很接近.

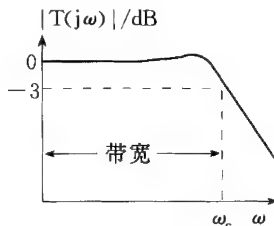
截止频率(cut-off frequency) 表示闭环系统频率响应特性的一种指标.系统闭环频率响应的幅值下降到零频率幅值以下3dB时所对应的频率ω_c称为截止频率,$0\sim\omega_c$间的频率范围称为频带宽度,常简称带宽.下图给出在伯德图上所示的截止频率和频带宽度.若系统闭环传递函数为$T(s)$,则有

$$20\lg|T(j\omega)|<20\lg|T(j0)|-3(\text{dB})$$
$$(\forall\omega>\omega_c).$$

频带宽度表征控制系统的响应速度.设计控制系统时,对频带宽度有如下要求:

1. 具有重现输入信号的能力,大的带宽对应于小的上升时间,或快速的响应.

2. 满足对高频噪声过滤特性的要求.



表示截止频率和频带宽度的伯德图

通常,这两个要求是互相冲突的,只能在两者之间进行折衷考虑.

带宽(band-width) 见"截止频率".

米哈依洛夫稳定判据(Michailov stability criterion) 根据特征曲线来判断线性定常系统稳定性的一种准则.它的理论基础是幅角原理.考虑闭环系统的特征方程式

$$D(s)=a_0s^n+a_1s^{n-1}+\cdots+a_n=0.$$

令$s=j\omega$,当ω由$-\infty$变到∞时,向量$D(j\omega)$端点在复平面上的轨迹称为$D(j\omega)$的特征曲线,又称为米哈依洛夫曲线.

幅角原理说明:当ω由$-\infty$变到∞时,$D(j\omega)$幅角的变化值等于$D(s)=0$的位于左半开s平面的根的数目$(n-m)$与位于右半闭s平面的根的数目m之差乘上π.闭环系统稳定的充分必要条件是:特征多项式所有的根都在左半开s平面.相应地,幅角变化由下式表示:

$$\Delta\arg D(j\omega)=n\pi\ (-\infty\leqslant\omega\leqslant\infty).$$

据此,米哈依洛夫稳定判据可表述为:闭环系统的特征多项式为$D(s)$,若向量$D(j\omega)$的幅角变化值在ω由$-\infty$变到∞时等于$n\pi$,其中n是$D(s)$的次数,则闭环系统是稳定的.通常,此判据也可表述为:当ω由0变到∞时,向量$D(j\omega)$转过角度$n\pi/2$,则闭环系统为稳定;或当ω由0变到∞时,特征曲线由正实轴开始,沿逆时针方向依次通过n个象限,则闭环系统为稳定.当特征曲线穿越坐标原点,同时转过$(n-1)$

个象限，那么系统处于稳定边缘.

奈奎斯特稳定判据(Nyquist stability criterion) 根据开环奈奎斯特图判断闭环系统稳定性的一种准则. 这一准则的理论基础是复变函数理论中的辐角原理. 设 $F(s)$ 为两个 s 的多项式之比，P 为 $F(s)$ 的极点数，Z 为 $F(s)$ 的零点数，它们位于 s 复平面上某一封闭曲线内，在此已考虑了多重极点和零点的情形. 设封闭曲线不通过 $F(s)$ 的任何极点或零点，则 s 平面上的这一封闭曲线映射到 $F(s)$ 复平面上也是一封闭曲线. 幅角原理指出：当 s 顺时针经过 s 平面上的整个封闭曲线时，轨迹 $F(s)$ 在 $F(s)$ 平面上顺时针包围原点的总次数 N 等于 $Z-P$. 利用幅角原理，研究图 1 所示的闭环系统，其开环传递函数为 $G(s)H(s)$，并令

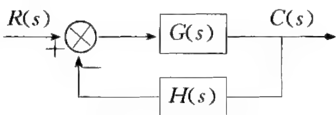


图 1　闭环系统

$$F(s)=1+G(s)H(s).$$

考虑到曲线 $1+G(j\omega)H(j\omega)$ 对原点的包围，等效于曲线 $G(j\omega)H(j\omega)$ 对 $-1+j0$ 点的包围，从而奈奎斯特稳定判据可表述为：对于 $G(s)H(s)$ 在 $j\omega$ 轴上既无极点又无零点的情况，如果开环传递函数 $G(s)H(s)$ 在右半 s 平面上有 k 个极点，并且

$$\lim_{s\to\infty} G(s)H(s)=\text{常量},$$

则为使闭环系统稳定，当 ω 从 $-\infty$ 变到 $+\infty$ 时，$G(j\omega)H(j\omega)$ 的轨迹必须反时针包围 $-1+j0$ 点 k 次. 对于 $G(s)H(s)$ 在 $j\omega$ 轴上有极点和(或)零点的情况，如果开环传递函数 $G(s)H(s)$ 在右半 s 平面上有 k 个极点，则为使系统稳定，当 s 顺时针通过修正后的奈奎斯特包线时，$G(j\omega)H(j\omega)$ 轨迹必须反时针方向包围 $-1+j0$ 点 k 次. 这里修正后的奈奎斯特包线就是采用无限小半径 ε 的半圆、避开虚轴上的零

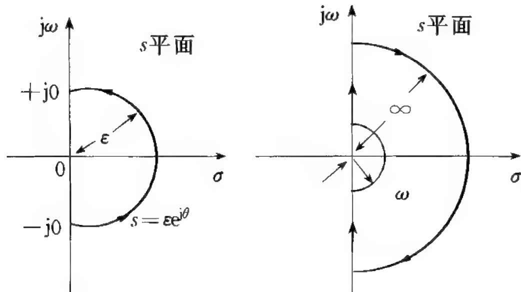


图2　修正后的奈奎斯特包线

极点的奈奎斯特包线. 图 2 表示避开原点的零极点的修正后的奈奎斯特包线. 应用奈奎斯特稳定判据来判断线性系统的稳定性时，$G(j\omega)H(j\omega)$ 轨迹有如下三种可能情况：

1. 不包围 $-1+j0$ 点. 如果 $G(s)H(s)$ 在右半 s 平面上没有极点，则闭环系统是稳定的，否则系统为不稳定.

2. 逆时针包围 $-1+j0$ 点. 如果逆时针包围的次数等于 $G(s)H(s)$ 在右半 s 平面上的极点数，则系统是稳定的，否则系统为不稳定.

3. 顺时针包围 $-1+j0$ 点. 不管 $G(s)H(s)$ 在右半 s 平面上是否有极点，系统都是不稳定的.

奈奎斯特稳定判据有如下特点：

1. 应用开环频率特性判断闭环系统稳定性，开环频率特性可由开环传递函数获得，也可应用实验数据绘制.

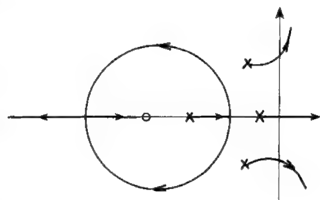
2. 便于研究系统参数和结构的改变对闭环系统稳定性的影响.

3. 便于研究包含延迟环节系统的稳定性.

4. 适当推广后，可用来分析某些非线性系统的稳定性.

奈奎斯特稳定判据是奈奎斯特(Nyquist，H.)于 1932 年提出的.

根轨迹法(root locus method) 确定闭环系统极点随参数变化情况的一种图解方法. 是以系统开环增益为参数，根据系统开环极点和零点，通过图解方法找出闭环系统极点随参数变化的轨迹. 本质上是一种通过图解来寻找特征方程根的方法，故因此得名. 如图为根轨迹的一个典型例子，图中×表示开环极点，◦表示开环零点，轨迹箭头展示了当增益由 0 变化到∞时闭环极点位置的变化方向. 一旦在复平面上标出开环极点和零点，利用作根轨迹图的一些规则，就可比较简单地绘制出根轨迹图. 由于闭环系统的瞬态响应行为主要由系统闭环极点所决定，而从根轨迹图上可以直观地看出每个开环极点或零点对闭环极点位置的影响，同时还指明了开环极点和零点应当怎样改变以使闭环极点位于复平面上的期望位置上，因而根轨迹法在线性控制系统的分析和综合中得到了广泛的应用. 采用这种方法的优点是可迅速地获得近似结果. 近年来，根轨迹法已被扩展到多变量系统.



根轨迹法是由埃文斯(Evans，W. R.)于 1948 年提出的.

控制系统的校正(correction of control systems) 经典控制理论中对控制系统进行调整的常用方法. 是通过引入附加装置来调整系统以使其满足给定瞬态和稳态性能要求. 对于大多数控制系统，仅仅调整增益并不能保证满足给定的性能指标，必须引入附加装置即校正装置来补偿原系统的缺陷. 校正的基本形式可分为串联校正(图 1)和并联校正

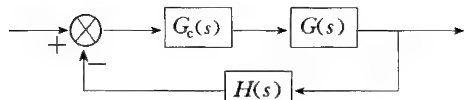


图1

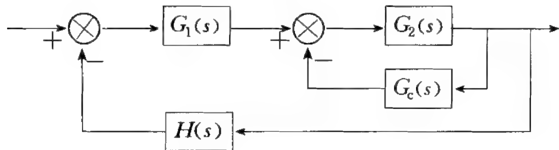


图2

(图 2). 在串联校正中，校正装置 $G_c(s)$ 与系统固有部分 $G(s)$ 构成串联连结. 在并联校正中，校正装置 $G_c(s)$ 被置于系统的一个内反馈回路中. 根据给定性能指标来决定相应的校正装置的传递函数 $G_c(s)$ 是线性控制系统综合中的基本问题，常用的方法有频率响应法和根轨迹法. 校正装置的基本类型有超前校正、滞后校正和滞后-超前校正，并常通过串联校正来实现. 校正装置可以采用电气的、机械的、气动的、液压的或混合形式的元件来构成.

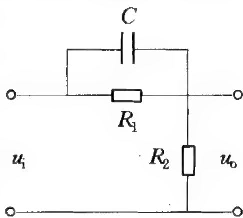
串联校正(tandem correction)　控制系统的一种基本校正方式(参见“控制系统的校正”).

并联校正(parallel correction)　控制系统的一种基本校正方式(参见“控制系统的校正”).

超前校正(lead correction)　控制系统的一种校正方式. 是校正装置的传递函数为

$$G_c(s) = \frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\alpha T}} \quad (\alpha < 1)$$

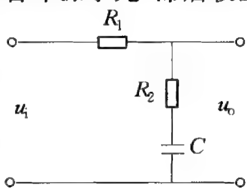
的一类串联校正. 在超前校正装置上输入一个正弦信号，则其输出量也是一个正弦信号，但在相位上超前于输入信号一个角度，超前校正之名即源于此. 在复平面上，超前校正装置的特点是其传递函数的零点总是位于极点的右方. 超前校正装置基本上是一个高通滤波器，主要作用是能使控制系统的瞬态响应得到显著改善，但不能显著改善稳态精度. 同时，如果存在噪声，则引入超前校正的结果会降低控制系统的信噪比，图中为用电阻、电容元件构成的一个超前校正网络.



滞后校正(lag correction)　控制系统的一种校正方式. 是校正装置的传递函数为

$$G_c(s) = \frac{1}{\beta}\left(\frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\beta T}}\right) \quad (\beta > 1)$$

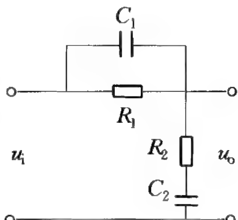
的一类串联校正. 当输入为正弦信号时，滞后校正装置的输出也是一个正弦信号，但在相位上落后于输入信号一个角度，滞后校正之名即源于此. 滞后校正装置的传递函数在复平面上的特点是极点总是位于零点的右边. 滞后校正装置基本上是一个低通滤波器，能使稳态精度得到显著改善，但瞬态响应的时间随之增加. 图中为用电阻、电容元件构成的一个滞后校正网络.



滞后-超前校正(lag-lead correction)　控制系统的一种校正方式. 是校正装置的传递函数为

$$G_c(s) = \left(\frac{s + \dfrac{1}{T_1}}{s + \dfrac{\beta}{T_1}}\right)\left(\frac{s + \dfrac{1}{T_2}}{s + \dfrac{1}{\beta T_2}}\right) \quad (\beta > 1)$$

的一类串联校正. 滞后-超前校正装置的传递函数中，第一项产生超前校正的作用，第二项产生滞后校正的作用. 当输入为正弦信号时，此装置的输出是带有相移的正弦信号，输出信号的相移性质与输入信号的频率有关. 当频率低时，相移是滞后的；当频率高时，相移是超前的. 滞后-超前校正综合了滞后校正和超前校正的优点，可同时改善系统的瞬态性能和稳态性能，即兼有快速响应特性和良好稳态精度. 图中为用电阻、电容元件构成的一个滞后-超前校正网络.



PID 调节(PID regulating)　经典控制理论中控制系统的一种基本调节方式. 是具有比例、积分和微分作用的一种线性调节规律. PID 调节的作用是将给定值 r 与被控变量的实际量测值 y 的偏差 $e=r-y$ 的比例、积分和微分信号综合成控制量来对被控过程进行控制. 这一控制量的表达式为

$$u = K_p e + \frac{1}{T_i}\int_0^t e\mathrm{d}t + T_d\frac{\mathrm{d}e}{\mathrm{d}t},$$

其中 K_p 为比例系数，T_i 为积分系数，T_d 为微分系数. 加大比例系数 K_p 可以减小系统的静差，但当 K_p 过大时，会使系统的动态品质变坏，引起被控量振荡，甚至导致闭环系统不稳定. 积分系数 T_i 大说明积分作用弱，反之则说明积分作用强. 增大 T_i 将减慢消除静差的过程，但可以减小超调，提高稳定性. 微分系数 T_d 增大，则微分作用加强，有助于减少超调，克服振荡，使系统趋于稳定，加快系统的响应速度，减小调整时间，从而改善系统的动态性能. 有众多的模拟的或数字的 PID 调节器产品可供选用. 用户只要根据实际应用环境合理地选择 K_p，T_i，T_d 三

个参数，使被控系统品质满足要求，选择参数的过程称为整定. 在工程实践中，模拟 PID 调节器的实用整定法有经验法、临界比例度法、衰减曲线法和响应曲线法. 数字 PID 调节器的实用整定法有扩展临界比例度法、扩展响应曲线法. 这些方法的本质是通过一系列在线实验，逐步地了解被控对象或环境的某些特征，根据经验或控制理论所揭示的基本关系来确定 K_p, T_i 和 T_d 三参数的适当取值.

复合控制(composite control)　一种系统控制方式. 是同时采用闭环控制和开环控制的一种控制

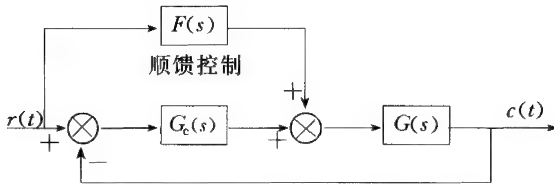


图 1

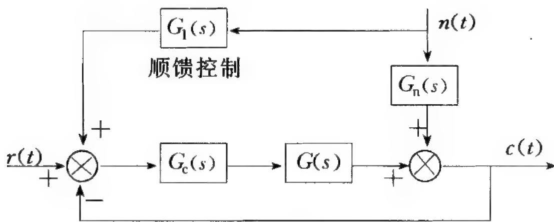
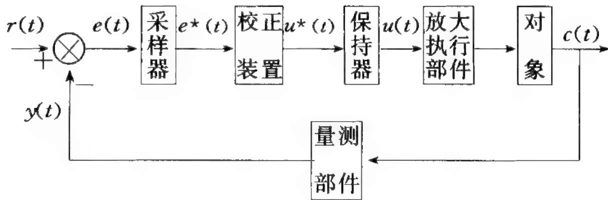


图 2

方式. 复合控制的组成结构，按其开环信号可分成为对输入的复合控制和对扰动的复合控制，分别如图 1 和图 2 所示. 图中 $G(s)$表征对象的特性，$G_c(s)$为校正装置的特性，$G_n(s)$为扰动对输出通道的特性，$F(s)$和$G_1(s)$分别为附加控制器的特性，$r(t)$，$n(t)$和$c(t)$分别为输入、扰动和输出. 在复合控制系统中所引入的开环控制常称为顺馈控制，置于开环通道的控制器称为顺馈控制器. 对于图 1 的系统，顺馈控制作用可保证输出量能快速跟随输入信号. 对于图 2 的系统，顺馈控制作用能使系统输出量在稳态下不受扰动的影响，即实现对扰动的静态不变性. 但是，在复合控制系统中，闭环控制仍是不可缺少的，由其他原因(如参数变动、其他扰动等)所引起的误差还需依靠反馈控制作用来消除. 理论分析和工程应用表明，采用复合控制可显著改善控制系统的动态和静态性能.

顺馈控制(feedforward control)　亦称前馈控制. 见“复合控制”.

采样控制理论(sampled-data control theory)　研究采样控制系统的分析和设计方法的理论. 采样控制理论的特点是系统中一处或几处信号具有脉冲序列或数字序列的形式，上图为一种典型的采样控制系统的组成原理框图. 通过引入采样器将误差信



号 $e(t)$转换为周期脉冲序列 $e^*(t)$，而经校正后的周期脉冲序列 $u^*(t)$再经保持器复原为分段连续信号 $u(t)$. 这样既可采用数字式装置作为校正装置以提高控制精度，又可避免 $u^*(t)$中的高频分量对系统的连续信号部分构成干扰. 在采样控制理论中，主要采用频率域方法，它是以 Z 变换为数学基础的. 描述采样控制系统基本特性的是脉冲传递函数，对采样控制系统的分析和设计都是建立在这一基本特性基础之上的. 采样控制理论的主要内容包括稳定性分析、瞬态响应、系统的校正原理和方法等.

采样器(sampler)　采样控制系统的基本元件之一. 是按一定采样方式把时间连续变化信号$x(t)$转化为相应的时间离散化信号$x^*(t)$的一类装置. 通常由开关元件构成，且采用周期采样方式. 图示为周期采样器的原理图，采样方式是等间隔的，每隔 T 秒钟开关闭合一次，使输入信号通过，其余时间信号被阻止通过，输出值为零. 一般地，采样持续时间与采样间隔相比常可忽略不计，从而采样器的输出-输入关系式为

$$x^*(t)=\begin{cases}x(kT) & (t=kT),\\ 0 & (t\neq kT)\end{cases}$$

$$(k=0,1,2,\cdots).$$

在采样控制系统的分析中，为了数学处理上的方便，常将上述关系式等价地表示为

$$x^*(t)=\sum_{k=0}^{\infty}x(kT)\delta(t-kT),$$

式中 $\delta(t-kT)$是作用点为 kT 的单位脉冲函数，且脉冲串中脉冲强度分别等于各采样瞬时上的采样值.

保持器(holder)　采样控制系统的基本元件之一. 是能将离散化信号转换为分段连续信号的一类装置. 在采样控制系统中，时间离散化信号是一个等间隔的脉冲串，脉冲作用点时间称为采样瞬时. 保持器的作用是按某种方式来构造两个相邻采样瞬时之间的信号，当这一信号为一个 n 阶的多项式函数时相应地称为 n 阶保持器. 控制工程中，常采用零阶保持器，即采用零阶多项式也就是常量来表述相邻采样瞬时之间的信号. 用$x(kT)$和 $x(t)$分别表示零阶保持器的输入和输出信号，则有

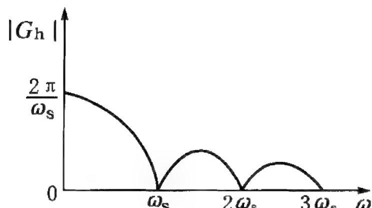
$$x(t) = x(kT),\ kT \leqslant t < (k+1)T,$$

式中 T 为信号的周期，$k=0,1,2,\cdots$零阶保持器的传递函数为

$$G_h(s) = \frac{1-e^{-Ts}}{s};$$

频率响应为

$$G_h(j\omega)=\frac{2\pi}{\omega_s}\frac{\sin\left(\frac{\pi\omega}{\omega_s}\right)}{\frac{\pi\omega}{\omega_s}}e^{-j\pi(\omega/\omega_s)}\quad\left(\omega_s=\frac{2\pi}{T}\right),$$



图示为其频率响应曲线.

Z 变换(Z transform)　用来处理离散函数的一种变换. 一个离散函数可以表示为

$$x^*(t) = \sum_{k=0}^{\infty}x(kT)\delta(t-kT)\quad(t \geqslant 0),$$

式中 T 为常数，$\delta(t-kT)$是作用点为 kT 的单位脉冲，$k=0,1,2,\cdots x^*(t)$的 Z 变换定义为

$$X(Z) = \sum_{k=0}^{\infty}x(kT)Z^{-k}.$$

相应地，Z 反变换的反演积分式为

$$x(kT) = \frac{1}{2\pi j}\oint X(Z)Z^{k-1}dZ.$$

根据 Z 变换的定义式，可建立起原函数 $x^*(t)$和象函数 $X(Z)$在函数上和在运算上的一一对应关系. Z 变换在求解差分方程中具有重要作用，可将差分方程变换为以 Z 为变量的代数方程，从而简化了求解过程. 在采样控制系统的分析和综合中，Z 变换是常用的方法之一，其作用与拉普拉斯变换在连续控制系统中所起的作用相似.

脉冲传递函数(pulse transfer function)　刻画线性采样系统的输出与输入间传递关系的一种表达式. 用 $X(Z)$和 $Y(Z)$分别表示系统的输入与输出的 Z 变换，脉冲传递函数 $G(Z)$定义为

$$G(Z)=\frac{Y(Z)}{X(Z)}.$$

对于图 1 所示的基本离散环节，其中 $G(s)$为连续部件的传递函数，T 为采样周期，则其脉冲传递函数为

图1

$$G(Z)=\sum_{k=0}^{\infty}g(kT)Z^{-k},$$

式中 $g(t)=\mathscr{L}^{-1}[G(s)]$，即 $G(s)$的拉普拉斯反变换，$g(kT)=g(t)|_{t=kT}$. 对于图 2 所示的两个基本离散环节的串联，在由 $G_1(s)$和 $G_2(s)$定出相应的 $G_1(Z)$和 $G_2(Z)$后，可得串联系统的脉冲传递函数为

$$G_T(Z) = G_2(Z)G_1(Z).$$

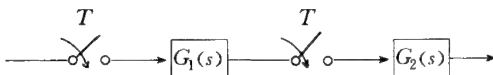


图2

对于闭环离散系统，利用上述计算规则，可容易地定出其输出-输入关系. 例如，图 3 和图 4 所示的闭环

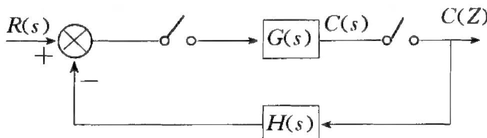


图3

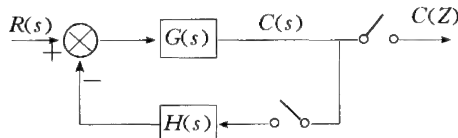


图4

离散系统的输出量 $C(Z)$分别可定义为

$$C(Z) = \frac{G(Z)R(Z)}{1+H(Z)G(Z)}$$

和

$$C(Z) = \frac{G(Z)R(Z)}{1+G(Z)H(Z)},$$

式中 $G(Z)H(Z)$是由 $G(s)H(s)$对应求得的脉冲传递函数，$G(Z)R(Z)$是由 $G(s)R(s)$对应求得的 Z 变换函数. 根据脉冲传递函数，可以方便地分析采样控制系统即闭环离散系统的稳定性. 例如，对于图 3 所示的闭环系统，其稳定性归结为判断特征方程

$$1 + H(Z)G(Z) = 0$$

的根 Z_i 是否全部位于复平面上的单位圆内，即是否成立 $|Z_i|<1(i=1,2,\cdots,n;n$ 为特征方程的次数). 当上式成立时，系统是稳定的，否则是不稳定的.

舒尔稳定(Schur stability)　关于离散线性系统的稳定性定义. 对于离散线性系统的特征多项式

$$F(z) = \sum_{i=0}^{n}a_iz^i\quad(a_n > 0),$$

如果其根均位于单位圆内，则称此系统是舒尔稳定的. 该定义是舒尔(Schur, I.)提出的.

舒尔-科恩稳定性判据(Schur-Cohn stability criterion)　判别舒尔稳定性的准则. 对于离散线性系统的特征多项式

$$F(z) = \sum_{i=0}^{n}a_iz^i\quad(a_n > 0),$$

其舒尔-科恩稳定性判据如下：

1. 对于多项式 $F(z)$，构造 $n\times n$ 对称矩阵 $C=[C_{ij}]$ 如下：

$$C_{ij}=\sum_{k=1}^{\min(i,j)}(a_{n-i+k}a_{n-j+k}-a_{i-k}a_{j-k}),$$

例如，当 $n=3$ 时，

$$C=\begin{bmatrix} a_3^2-a_0^2 & a_2a_3-a_1a_0 & a_1a_3-a_2a_0 \\ a_2a_3-a_1a_0 & a_3^2+a_2^2-a_0^2-a_1^2 & a_2a_3-a_1a_0 \\ a_1a_3-a_2a_0 & a_2a_3-a_1a_0 & a_3^2-a_0^2 \end{bmatrix},$$

舒尔稳定的充分必要条件是 C 为正定矩阵.

2. 对于多项式 $F(z)$，构造两个 $(n-1)$ 阶方阵如下：

$$A_{n-1}^1=\begin{bmatrix} a_n & a_{n-1} & a_{n-2} & \cdots & a_2 \\ 0 & a_n & a_{n-1} & \cdots & a_3 \\ 0 & 0 & a_n & \cdots & a_4 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_n \end{bmatrix},$$

$$A_{n-1}^2=\begin{bmatrix} 0 & 0 & 0 & \cdots & a_0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & a_0 & \cdots & a_{n-4} \\ 0 & a_0 & a_1 & \cdots & a_{n-3} \\ a_0 & a_1 & a_2 & \cdots & a_{n-2} \end{bmatrix},$$

则 $F(z)$ 是舒尔稳定的充分必要条件是

$$F(1)>0,\ (-1)^nF(-1)>0,$$

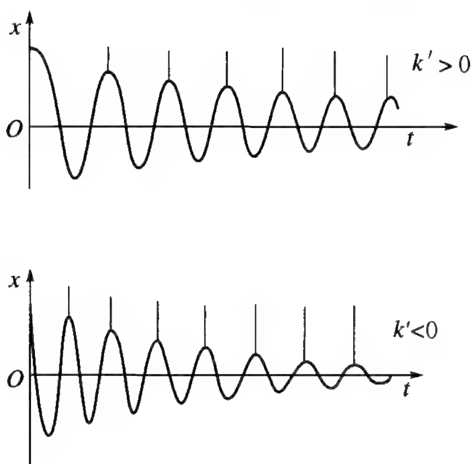
且 $A_{n-1}^1\pm A_{n-1}^2$ 的所有内矩阵的行列式均大于零，这里矩阵的内矩阵是指将原矩阵的前 i 行、最后 i 行和前 i 列、最后 i 列去掉之后形成的矩阵($i=0,1,2,\cdots$).

非线性现象(nonlinear phenomena)　只能出现于非线性系统中的某些特殊动力学行为. 非线性现象是非线性系统中独有的反映其运动本质的一类现象，在线性系统中是见不到的，也不能用线性系统理论来解释. 常见的比较重要的非线性现象有非线性振荡、多值响应和跳跃谐振、自激振荡、频率捕捉、异步抑制、分岔、混沌等. 对非线性现象的研究是非线性系统理论的一项基本内容.

非线性振荡(nonlinear oscillation)　一类出现于非线性系统中的区别于线性振荡的振荡现象. 非线性振荡的基本特性之一是频率对振幅的依赖性. 典型例子为作自由振荡的达芬(Duffin, R. J.)系统，其位移 x 的运动方程为

$$m\ddot{x}+f\dot{x}+kx+k'x^3=0,$$

其中 $\dot{x}$ 和 $\ddot{x}$ 为位移 x 的一阶和二阶导数，$m>0$ 为质量，$f>0$ 为摩擦系数，$k>0$ 和 k' 为弹簧系数，$k'>0$ 对应于硬弹簧，$k'<0$ 对应于软弹簧. 系统受非零初始条件作用下的自由振荡如下图所示.

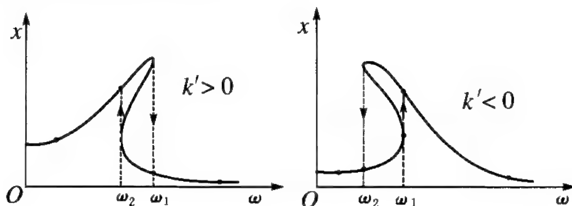


结果显示，振荡频率依赖于振荡振幅，$k'>0$ 时频率随振幅减小而减小，$k'<0$ 时频率随振幅减小而增加. 非线性振荡的另一基本特性是分谐波振荡现象. 分谐波振荡是一种非线性稳态振荡，它的发生取决于系统参数、初始条件以及外作用函数的振幅和频率，同时必须出现有某种冲击，如突然改变外作用函数的振幅和频率. 分谐波振荡一旦被激发，其现象表现为其频率恰等于 ω/n，这里 ω 是外作用频率，正整数 n 称为分谐波振荡的阶数. 分谐波振荡在某个频率范围内可能是非常稳定的，即不管外作用函数的频率如何改变，稳态振荡的频率始终为 ω/n. 但如果外作用频率改变到一个新的值 ω'，那么分谐波振荡不是消失，就是频率也改变到 ω'/n.

多值响应(multi-value response)　出现于某些非线性系统强迫振荡中的非线性现象. 其特征是系统响应振幅在某些输入频率值时会发生跳跃，而在某个输入频率段内可具有两个不同振幅值的稳定振荡. 一个典型的例子是处于强迫振荡的达芬(Duffin, R. J.)系统，其运动方程为

$$m\ddot{x}+f\dot{x}+kx+k'x^3=P\cos\omega t,$$

其中 P 为常值，$\dot{x}$ 和 $\ddot{x}$ 为位移 x 的一阶和二阶导数，m 为质量，$k>0$ 为弹簧系数，阻尼 f 很小，外作用振幅 P 足够大，弹性系数 $k'>0$ 和 $k'<0$ 分别对应于硬弹簧和软弹簧.



系统在强迫振荡下观察到的响应振幅 x 和输入频率 ω 间的关系特性如上图所示. 实验结果显示，当 ω 增加到 ω_1 再继续增加或 ω 减少到 ω_2 再继续减小时，会出现振幅值的跳跃，在 ω_1 和 ω_2 的频率段内系统具有两个可能的稳定振荡，即跳跃谐振和多值响应.

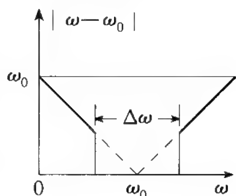
跳跃谐振(jumping resonance) 见"多值响应".

自激振荡(self-exciting oscillation) 出现于某些非线性系统中的一种自由振荡.一个典型例子是范达波尔(Van der Pol)方程所描述的系统,方程形式为

$$m\ddot{x}-f(1-x^2)\dot{x}-kx=0\ (m>0,\ f>0,\ k>0).$$

其中 $\dot{x}$ 和 $\ddot{x}$ 为变量 x 的一阶和二阶导数.分析表明:当 x 的值很小时,阻尼 f 是负的,因而运动发散;当 x 的值很大时,阻尼 f 是正的,因而运动衰减.所以,不管由什么初始条件出发,系统运动都趋向于一个持续振荡,即自激振荡.对于更为复杂的系统,可以具有两个或多个不同振幅的自激振荡,处于不同区域的初始条件出发的运动可能趋于不同振幅的自激振荡.在相平面分析中,这种二阶系统的自激振荡即为极限环.

频率捕捉(frequency catching) 一类非线性现象.出现于某些非线性系统中的、自激振荡频率被同步于外作用频率的一类非线性现象.频率捕捉的具体表现是:对一个能出现频率为 ω_0 的自激振荡的系统,若加入一个频率为 ω 的周期性外作用,就会产生差拍现象.但当两个频率之差落入到某个范围 $\Delta\omega$ 内时,差拍消失,自激振荡频率 ω_0 为外作用频率 ω 所同步,也即 ω_0 为 ω 所捕捉.发生频率捕捉现象的频带段称为捕捉区.表征频率捕捉现象的特性如图所示,其中捕捉区内的虚线部分反映了线性系统的情况,它表明对于线性系统只有当 $\omega=\omega_0$ 的一个点上才能达到 $|\omega-\omega_0|=0$,而对于一类非线性系统可在捕捉区内的每个 ω 值上达到 $|\omega-\omega_0|=0$.



异步抑制(asynchronous restraining) 一类用以抑制自激振荡频率的现象.指在一些具有频率为 ω_0 的自激振荡的非线性系统中,通过引入外作用来使系统处于另一频率 ω_1 的强制振荡状态,以抑制自激振荡频率的一类现象.异步抑制在工程中有着很多应用,并且,在异步抑制中,两个频率值 ω_1 和 ω_0 是互不关联的.

输入-输出稳定性(input-output stability) 也称为有界输入-有界输出稳定性(简称 BIBO 稳定性).系统的一种稳定性.是从系统的外部特性来考察的系统稳定性行为.设 t_0 为初始时刻,若对零初始条件,每一个定义在$[t_0,\infty)$上的有界输入产生一个在$[t_0,\infty)$上的有界输出,则称系统是输入-输出稳定的.

描述函数(describing function) 一种表征元件特性的函数.是一种用一次谐波响应近似表征非线性元件特性的函数.假设非线性特性是对称于原点的,X 表示零相位正弦输入的振幅,Y_1 表示输出信号中一次谐波分量的振幅,φ_1 表示输出信号一次谐波分量的相位移,那么非线性元件的描述函数 N 定义为

$$N=\frac{Y_1}{X}\arg\varphi_1,$$

对于不包含储能元件的非线性元件,N 只是输入振幅 X 的函数.作为典型例子,对于图示的饱和非线性(图 1)和滞环非线性(图 2),它们的描述函数分别为

$$N=\frac{2k}{\pi}\left[\arcsin\left(\frac{a}{X}+\frac{a}{X}\sqrt{1-\frac{a}{X}^2}\right)\right]$$

和

$$N=\frac{4M}{\pi X}\cdot\left[-\arcsin\frac{h}{X}\right].$$

描述函数在分析非线性系统的自激振荡中有着重要的应用.

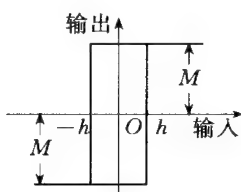


图1

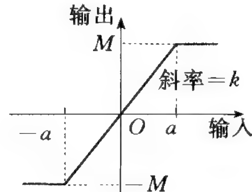


图2

描述函数法(describing function method) 分析系统自激振荡及稳定性的方法.指基于描述函数的用以分析非线性控制系统自激振荡及其稳定性的一种近似方法.在描述函数法中,所研究的非线性控制系统如图 1 所示,其中线性部分设为是最小相位的且用频率响应 $G(\mathrm{j}\omega)$ 来表征,非线性元件用描述函数 N 来表征.为了分析系统的稳定性,在复平面上同时画出 $-1/N$ 随振幅 $X=0\to\infty$ 和 $G(\mathrm{j}\omega)$ 随频率 $\omega=0\to\infty$ 的轨迹,并且有结论:如果 $-1/N$轨迹没有被 $G(\mathrm{j}\omega)$ 轨迹所包围,则系统为稳定的,即稳态时没有自激振荡;如果 $-1/N$ 轨迹被 $G(\mathrm{j}\omega)$轨迹所包围,则系统为不稳定;如图 2 所示,若 $-1/N$ 轨迹和$G(\mathrm{j}\omega)$轨迹相交,那么系统的输出可能出现自激振荡.进一步,$-1/N$ 轨迹沿 X 增加方向越出 $G(\mathrm{j}\omega)$轨迹包围的交点,则(如图 2 中 B 点),则对应于稳定的自激振荡;反之进入$G(\mathrm{j}\omega)$轨迹包围的交点(如图 2 中 A 点),则对应于不稳定的自激

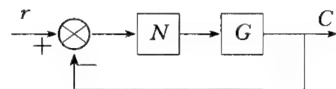


图1

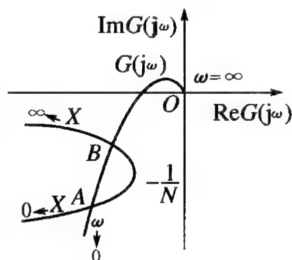


图 2

振荡.只有稳定的自激振荡才是实际上可以出现的，其振幅和频率即为相应交点处 $-1/N$ 轨迹上的 X 值和 $G(j\omega)$ 轨迹上的 ω 值.一般地，只有当系统的线性部分对高次谐波分量具有较好的衰减特性时，描述函数法才具有较好准确度.早期文献中也称描述函数法为谐波平衡法.

谐波平衡法(harmonic balance method)　对描述函数法的另一种称谓.见“描述函数法”.

相平面法(phase-plane method)　一种在二维平面上定性分析二阶系统运动特性的图示方法.对于由微分方程 $\ddot{x}+f(x,\dot{x})=0$($f(x,\dot{x})$可为线性或非线性函数)所描述的二阶系统，其运动过程常可以以时间 t 作为参变量，采用 $\dot{x}(t)$ 和 $x(t)$ 的关系图来表示.在以 x 和其一阶导数 $\dot{x}$ 为直角坐标的相平面上，确定的初始条件 $\dot{x}(0)$ 和 $x(0)$ 对应于相平面上的一个初始点；而随着时间 t 的变化，运动过程被表征为由初始点出发在相平面上描绘出的一条曲线，称为相轨迹.对于所有可能的初始条件，运动过程由一族相轨迹来表征.相平面上对应于同时满足 $\dot{x}=0$ 和 $f(x,\dot{x})=0$ 的点称为奇点，奇点也就是运动的平衡点.当微分方程满足解的惟一性条件时，除了在奇点处以外，由所有可能初始条件决定的相轨迹都不会彼此相交.当系统微分方程是简单的或分段线性的方程时，可采用解析法来做出相轨迹；而当系统微分方程难以甚至不可能用解析法求解时，常采用图解法来做出相轨迹.采用相平面分析，使人们能直观地看到任何可能初始条件下的全部运动过程，并可简单地判断平衡点的稳定性以及自激振荡的存在性和稳定性.在相轨迹上没有直接显示时间信息，但若有必要，通过简单的计算也可在其上进行时间标度.在处理严重非线性的系统时，相平面法有着特别的重要性.相平面法是由庞加莱(Poincaré,(J.-)H.)首先提出的，是研究二阶微分方程的定性行为(稳定性、极限环、奇点类型等)的重要手段之一.事实上，相平面法是状态空间法的一类特殊情况.相平面法的局限性在于通常只适用于二阶系统.

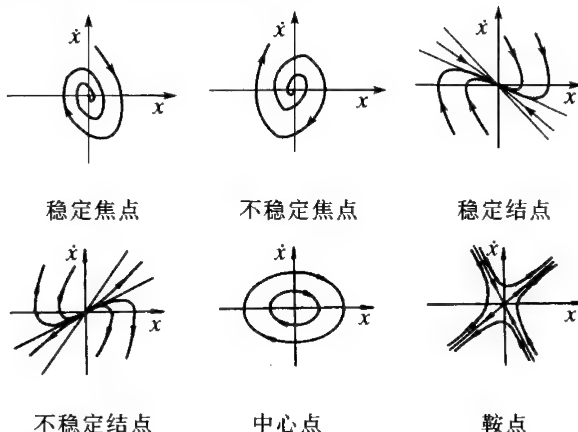
奇点(singular point)　二阶微分方程解轨线上的一类特殊点.它是由微分方程 $\ddot{x}+f(x,\dot{x})=0$ ($f(x,\dot{x})$可为线性或非线性函数)所描述的二阶系统，在以 x 和其一阶导数 $\dot{x}$ 为直角坐标的相平面上，同时满足 $\dot{x}=0$ 和 $f(x,\dot{x})=0$ 的点.奇点对应于系统运动过程的平衡点.由所有可能初始条件决定的相轨迹最终达到或者逼近某个奇点，或者围绕某个奇点，或者离开某个奇点，它们分别对应于该平衡点是李亚普诺夫(Liapunov)意义下渐近稳定的、稳定的、不稳定的.对于二阶线性定常系统

$$\ddot{x}+a\dot{x}+bx=0,$$

其中 a 和 b 为常数且 $b\neq0$，系统的奇点位于相平面的原点.由此系统特征方程

$$s^2+as+b=0$$

的两个根 λ_1 和 λ_2 在复平面上的位置可完全确定其奇点的特性.图中是六种可能的奇点类型及相轨迹形状：当 λ_1 和 λ_2 为共轭复数且位于左半复平面内，奇点为稳定焦点；当 λ_1 和 λ_2 为共轭复数且位于右半复平面内，奇点为不稳定焦点；当 λ_1 和 λ_2 为实数且位于左半复平面内，奇点为稳定结点；当 λ_1 和 λ_2 为实数且位于右半复平面内，奇点为不稳定结点；当 λ_1 和 λ_2 为共轭虚数，奇点为中心点；当 λ_1 和 λ_2 为实数且分别位于左半和右半复平面内，奇点为鞍点.一般地，奇点的特性完全地刻画了奇点的稳定性及系统在奇点附近的运动过程形态.



极限环(limit-cycle)　二阶系统的孤立封闭轨线.由微分方程 $\ddot{x}+f(x,\dot{x})=0$($f(x,\dot{x})$为非线性函数)所描述的二阶系统，在以 x 及其一阶导数 $\dot{x}$ 为直角坐标的相平面上的一个孤立封闭解轨迹.对于不少的系统，可以有多个极限环.由微分方程解的惟一性可保证，极限环内部或外部的相轨迹不可能穿越极限环而进入其外部或内部.极限环可分为稳定极限环、不稳定极限环、半稳定极限环三种情况.

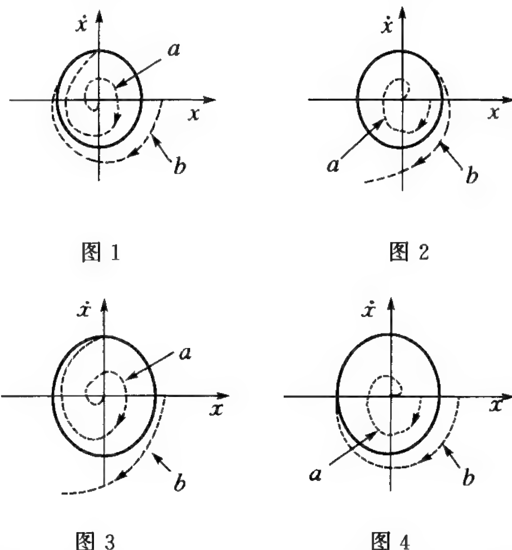


图 1为稳定极限环，特点是起始于极限环外部和内部附近区域的相轨迹均收敛于此极限环. 稳定极限环对应于系统中存在的自激振荡. 图 2 是不稳定极限环，特点是极限环附近的相轨迹均从此极限环发散出去. 相应于不稳定极限环的自激振荡实际上是不可能出现的. 图 3 和图 4 为半稳定极限环，特点是起始于极限环外部(或内部)的相轨迹从极限环发散出去，起始于极限环内部(或外部)的相轨迹收敛于极限环. 对应于半稳定极限环的自激振荡实际上也是观测不到的. 从保证控制系统的稳定性的角度考虑，对具有稳定极限环的系统应当尽量减小极限环，对具有不稳定极限环的系统应当尽可能增大极限环，而对具有半稳定极限环系统则只是对图 4 的情况才是允许的. 在某些工程装置中，也常希望将系统设计为具有期望参数(频率和振幅)的稳定极限环，以获得期望的自激振荡信号，例如钟表和各类无线电设备中的振荡信号源等.

分段线性化法(piecewise linearization method) 处理二阶非线性系统的一种近似方法. 它是通过对非线性特性进行分段线性化处理来分析非线性系统的一种方法. 利用分段线性化法分析系统时，先把非线性特性曲线划分成若干个区段且用直线段代替曲线段，对每一区段的等效线性化系统进行分析，再把各个区段的分析结果按时间顺序加以衔接，就得到非线性系统的近似分析结果. 对于非线性程度较低的系统，分段线性化法常具有较好的分析结果. 对于非线性程度高的系统，分段线性化法原则上仍可适用，但分析准确度取决于线性化的区段数的多少，常导致计算量的增加.

不变原理(invariance principle)　一种可减小和消除扰动对控制系统的不利影响的系统设计原理. 不变原理适用于扰动可直接量测的情况，其基本思想是利用基于扰动的控制作用来抵消扰动对系统输出的影响，因此，双通道原理构成了不变原理系统的一个基本结构原则，即要求扰动作用点和被控变量量测点之间至少存在两个信号传递通道. 这种按双通道原理建立的控制系统，形式上属于同时包含按偏差的闭环控制和按扰动的开环控制的复合控制系统. 不变原理建立了系统输出完全不受扰动影响和稳态时不受扰动影响的条件，为设计和构成高精度和高性能的自动控制系统提供了理论依据.

线 性 系 统

线性系统(linear systems)　系统过程满足叠加原理时的一种模式. 在这类系统中，当系统无控制作用时，其过程的和及常数倍均仍为系统的过程. 在有控制作用时，由初值与控制共同决定的过程可以分解为由具初值但零控制决定的过程与对应控制但零初值决定的过程之和. 在系统的数学描述方法上，这类系统常由一状态满足的线性方程附加一线性控制所构成，系统的输出也只线性地依赖于系统的状态与控制.

从用数学描述系统的类型分，线性系统有：用常系数线性微分方程描述的定常连续线性系统；用常系数差分迭代方程描述的定常离散线性系统；用变系数线性微分(差分)方程描述的时变连续(离散)线性系统；用线性偏微分方程和积分方程描述的分布参数线性系统；用线性随机微分方程描述的随机线性系统以及方程中含有状态的时间滞后量的时滞线性系统等.

定常连续线性系统是理论发展最为完善且应用广泛的一类系统，除用微分方程组描述的状态空间模式外，也有基于拉普拉斯变换的传递函数矩阵模式. 后者由于可借助频率特性进行讨论而为工程界所钟爱. 这类系统无论是结构性质(能控性、能观测性、能镇定性等)、稳定性及二次性能指标的最优控制均已有完整的理论，其研究工具主要是线性代数和与拉普拉斯变换有关的复变函数理论. 近年来主要关心的课题是为抑制系统中含有的不确定性和干扰而发展起来的线性鲁棒控制，包括 H_∞ 理论和具参数不确定性的鲁棒控制理论等.

时变线性系统由于研究工具的不足，虽然在理论上可以得到与定常线性系统相近的一系列结论，但无论是这些结论的判定还是应用，困难均很大. 比较有效的是基于时变线性方程解的形式进行估计或利用李亚普诺夫方法进行研究，其中一些只是近似的或有一定保守性的充分条件.

分布参数线性系统研究的主要对象是讨论无穷维线性系统，应用线性泛函分析工具可以从理论上讨论一些基本问题. 但由于在无穷维与有穷维之间存在本质的区别，控制实际希望用有穷维甚至低阶系统来实现，这就为分析和设计无穷维系统带来大量的理论问题，例如控制溢出、观察溢出等.

由维纳(Wiener，N.)的预测和滤波及其后的卡尔曼滤波为开端发展起来的线性随机控制系统理论，主要讨论系统稳定性、系统辨识与参数估计问题以及自适应与自校正控制. 由于实际控制的需要，这一类问题的研究中主要是提出并实现各种有效的算法，并在理论上回答这些算法应具有的属性(例如收敛性、适定性等).

线性系统只是实际系统的一个近似模式. 在系统平衡点附近，当线性模式是渐近稳定或本质不稳定时，它的这类特征可以反映真实系统，但在远离平衡点作大范围运动时，线性系统模式就难以适应需要，特别对一些本质非线性的现象(自振、分岔、混沌

现象等)用线性模式讨论是不可能奏效的.

线性控制系统(linear control systems) 动态行为可用一线性数学模型表征的控制系统. 当时间变量连续时,它是由一向量常微分方程和一代数方程来描述:

$$\begin{cases} \dot{x} = A(t)x + B(t)u, \\ y = C(t)x. \end{cases}$$

当时间变量离散时,它是由一向量差分方程和一代数方程来描述:

$$x_{k+1} = \Phi_k x_k + \Gamma_k u_k,$$
$$y_k = H_k x_k,$$

其中 $x \in \mathrm{R}^n(x_k \in \mathrm{R}^n)$, $u \in \mathrm{R}^r(u_k \in \mathrm{R}^r)$, $y \in \mathrm{R}^m(y_k \in \mathrm{R}^m)$, $A(t)(\Phi_k)$, $B(t)(\Gamma_k)$, $C(t)(H_k)$为适当维数矩阵. 前者称为连续线性时变控制系统,简记为 $\Sigma(A(t),B(t),C(t))$, 后者称为离散线性时变控制系统,简记为 $\Sigma(\Phi_k,\Gamma_k,H_k)$. 值得注意的是:一个线性控制系统,特别是线性定常控制系统,可以有各种不同的描述方法. 这里是最常用的状态空间方法.

系统的状态向量(state vector of systems) 确定和了解系统情况的一些变量. 控制系统的状态变量是指为刻画该系统动态行为所需要的一组独立变量. 将状态变量排成一向量就是状态向量;输入向量又称控制向量,它是改变系统动态行为的主要手段,通常是由设计者来选择的;输出向量又称量测向量,它是人们了解系统动态行为信息的来源. 例如在连续线性时变控制系统

$$\begin{cases} \dot{x} = A(t)x + B(t)u, \\ y = C(t)x \end{cases} \tag{1}$$

中,$x \in \mathrm{R}^n$ 是状态向量,$u \in \mathrm{R}^r$ 是控制向量,$y \in \mathrm{R}^m$ 是输出向量. 当 $A(t)=A, B(t)=B, C(t)=C$ 时,则称(1)为连续线性定常系统. 对连续线性定常控制系统,为了提高控制和量测设备的利用率,通常设

$$\operatorname{rank} B = r,\ \operatorname{rank} C = m,$$

其中 rank(·)表示矩阵的秩.

系统的控制向量(control vector of systems) 见“系统的状态向量”.

系统的输出向量(output vector of systems) 见“系统的状态向量”.

状态方程(state equation) 刻画系统输入和状态关系的表达式. 状态向量所满足的向量常微分方程称为控制系统的状态方程. 如连续线性时变控制系统

$$\begin{cases} \dot{x} = A(t)x + B(t)u, & (1) \\ y = C(t)x & (2) \end{cases}$$

中的(1)式称为状态方程. 如果状态向量的初始条件 $x(t_0)=x_0$ 和 $t \geqslant t_0$ 时的输入都已知,则可从(1)式完全决定 $t \geqslant t_0$ 时刻的所有状态 $x(t)$,因而控制系统的动态行为就完全确定了. 刻画控制系统的输出与状态之联系的代数关系称为输出(或量测)方程. (2)式便是输出方程. 输出方程提供了人们通过量测数据了解系统状态变化的信息. 状态方程和输出方程是控制系统数学模型的重要组成部分.

输出方程(output equation) 见“状态方程”.

能控性(controllability) 线性控制系统理论的一个基本概念. 直观上讲,能控性刻画了控制输入对其状态变量的制约能力. 它反映的是系统的整体结构性质. 对给定的时变连续线性系统 $\Sigma(A(t),B(t),C(t))$, 如果对 t_0 时刻给定的状态 x_0,能找到一时刻 $t_1 > t_0$ 和确定在$[t_0,t_1]$上的控制 $u(t)$,使得系统 Σ 在 $u(t)$ 的控制下有 $x(t_1)=0$,则称状态 x_0 在 t_0 时刻为能控的;如果每一状态 $x_0 \in \mathrm{R}^n$ 在 t_0 时刻都是能控的,则称系统 Σ 在 t_0 时刻为能控的; 如果系统 Σ 在$[t_0,T]$ 的每个时刻都是能控的,则称系统 Σ 在$[t_0,T]$ 上完全能控,或简称能控. 能控性亦称为可控性.

能达性(reachability) 反映控制对其状态制约能力的基本概念. 给定线性系统 $\Sigma(A(t),B(t),C(t))$, 如果对 t_0 时刻给定的状态$x_1 \in \mathrm{R}^n$,存在一个 $t_1 > t_0$ 和定义在$[t_0,t_1]$上的控制$\bar{u}(t)$,使得系统 Σ 在 $\bar{u}(t)$ 的控制下从 t_0 时刻零初态出发的轨线有 $x(t_1)=x_1$,则称系统的状态 x_1 是 t_0 时刻能达的;如果每一个状态 $x_1 \in \mathrm{R}^n$ 都是 t_0 时刻能达的,则称系统 Σ 为 t_0 时刻能达的;如果系统 Σ 在$[t_0,T]$ 上的每一时刻都是能达的,则称系统 Σ 在$[t_0,T]$ 上 是能达的,或简称能达. 能达性亦称为可达性.

对线性连续系统,能控性和能达性是一致的. 它们都反映系统 Σ 的结构性质,而与系统的具体输入无关. 但对线性离散系统,只有当 Φ_k 皆为满秩阵时两者才是一致的.

能控性矩阵(controllability matrix) 亦称格拉姆矩阵. 线性系统能控性的数学表达. 连续线性时变控制系统 $\Sigma(A(t),B(t),C(t))$ 在 t_0 时刻的能控性矩阵为

$$W(t_1,t_0) \triangleq \int_{t_0}^{t_1} \Phi(t_1,\tau)B(\tau)B^{\mathrm{T}}(\tau)\Phi^{\mathrm{T}}(t_1,\tau)\mathrm{d}\tau$$
$$(t_1 > t_0),$$

其中 $\Phi(t,\tau)$是系统 Σ 的状态转移阵;而线性定常系统 $\Sigma(A,B,C)$ 的能控性可通过下述判别矩阵:

$$Q_c \triangleq [B\ AB\ A^2B\ \cdots\ A^{n-1}B]$$

来判定,其中 n 为方阵 A 的行(列)数.

能控性秩条件(controllability rank condition) 刻画系统能控性的数学条件. 系统 $\Sigma(A(t),B(t),C(t))$ 在 t_0 时刻(完全)能控的充分必要条件是:存在 $t_1(>t_0)$,使 Σ 的能控性矩阵 $W(t_1,t_0)$ 满秩;而线性定常系统 $\Sigma(A,B,C)$(完全) 能控的充分 必要条件是:它的判别矩阵 Q_c(行)满秩. 使得

$$\mathrm{rank}\,[B\ AB\ \cdots\ A^{\nu-1}B] = n$$

的最小整数 ν 称为系统 $\Sigma(A,B,C)$ 的能控指数.

线性系统能控的判别条件(criterion of controllability for linear systems) 判别线性系统能控性的准则. 给定连续线性时变系统 $\Sigma(A(t),B(t),C(t))$,下述命题等价:

1. $\Sigma(A(t),B(t),C(t))$ 在 t_0 时刻能控.

2. 存在某个时刻 $t_1(>t_0)$,使能控性矩阵

$$W(t_1,t_0) \triangleq \int_{t_0}^{t_1} \Phi(t_1,\tau)B(\tau)B^{\mathrm{T}}(\tau)\Phi^{\mathrm{T}}(t_1,\tau)\mathrm{d}\tau$$

为正定阵.

3. 存在某个时刻 $t_1(>t_0)$,使时变矩阵 $\Phi(t_1,\tau)\cdot B(\tau)$在$[t_1,t_0]$上行满秩.

这里 $\Phi(t_1,\tau)$是齐次线性方程 $\dot{Y}=A(t)Y$ 的基本解矩阵.

对线性定常系统 $\Sigma(A,B,C)$,下列命题等价:

1. 系统 $\Sigma(A,B,C)$ 能控.

2. 对任给的 $t_1(>t_0)$,有

$$W(t_1,t_0) \triangleq \int_{t_0}^{t_1} \mathrm{e}^{A(t_1-\tau)}BB^{\mathrm{T}}\mathrm{e}^{A^{\mathrm{T}}(t_1-\tau)}\mathrm{d}\tau > 0.$$

3. $\mathrm{rank}[B\ AB\ A^2B\ \cdots\ A^{n-1}B]=n$.

4. $\mathrm{rank}[\lambda I-A\ \ B]=n$,
$\forall \lambda \in \sigma(A) \triangleq \{\lambda | \det(\lambda I - A) = 0\}$.

5. $\bigcup_{j=1}^{n} R(A^{j-1}B)=\mathbb{R}^n$.

其中 $R(\cdot)$为算子的值域. 判据 5 是系统 Σ 能控性的一种几何形式判据.

能观测性(observability) 线性控制系统理论的一个基本概念. 它刻画了系统的输出对状态的判断能力. 给定线性系统$\Sigma(A(t),B(t),C(t))$,对 t_0 时刻给定的初态 x_0,如果存在 $t_1>t_0$,使得通过量测在$[t_0,t_1]$上的输出 $y(t)$和已知的控制输入 $u(t)$,能惟一决定出 $x(t_0)=x_0$,则称初态 x_0 为 t_0 时刻能观测的;如果每一个初态 $x_0\in\mathbb{R}^n$ 都是 t_0 时刻能观测的,则称系统 Σ 在 t_0 时刻为能观测的;如果系统 Σ 在$[t_0,T]$上的每个时刻都是能观测的,则称系统 Σ 在$[t_0,T]$上是完全能观测的,或简称能观. 能观测性亦称可观测性.

能检测性(detectability) 控制系统的一种结构性质. 给定线性定常系统 $\Sigma(A,B,C)$,如果存在常数阵 $G\in\mathbb{R}^{n\times m}$使得 $A+GC$ 的全部特征值 $\sigma(A+GC)\subset\mathbb{C}^-$,则称系统 $\Sigma(A,B,C)$是能检测的. 有时也称矩阵对(A,C)是能检测的. 系统 $\Sigma(A,B,C)$能检测的充分必要条件是系统 $\Sigma(A,B,C)$ 的不能观测极点都是稳定的. 系统 $\Sigma(A,B,C)$ 的能检测概念在为其设计状态观测器时有重要作用.

能观测性矩阵(observability matrix) 线性系统能观测性的数学表达. 连续线性时变系统 $\Sigma(A(t),B(t),C(t))$在时刻 t_0 的能观测性矩阵为

$$M(t_1,t_0) = \int_{t_0}^{t_1} \Phi^{\mathrm{T}}(\tau,t_0)C^{\mathrm{T}}(\tau)C(\tau)\Phi(\tau,t_0)\mathrm{d}\tau \quad (t_1 > t_0),$$

其中 $\Phi(t,\tau)$是相应齐次方程的基本解阵. 线性定常系统 $\Sigma(A,B,C)$的能观测矩阵为

$$\int_{t_0}^{t_1} \mathrm{e}^{A^{\mathrm{T}}\tau}C^{\mathrm{T}}C\mathrm{e}^{A\tau}\mathrm{d}\tau.$$

能观测性秩条件(observability rank condition) 刻画线性系统能观测性的数学条件. 连续线性时变系统 $\Sigma(A(t),B(t),C(t))$在时刻 t_0(完全)能观测的充分必要条件是:存在 $t_1 > t_0$ 使得其能观测矩阵 $M(t_1,t_0) > 0$;而 n 阶线性定常系统 $\Sigma(A,B,C)$ (完全)能观测的充分必要条件是:它的判别矩阵

$$R_0 = [C^{\mathrm{T}}\ A^{\mathrm{T}}C^{\mathrm{T}}\ \cdots\ (A^{\mathrm{T}})^{n-1}C^{\mathrm{T}}]$$

列满秩,即 $\mathrm{rank}R_0=n$. 使得

$$\mathrm{rank}[C^{\mathrm{T}}\ A^{\mathrm{T}}C^{\mathrm{T}}\ \cdots\ (A^{\mathrm{T}})^{\mu-1}C^{\mathrm{T}}] = n$$

成立的最小整数 μ 称为系统$\Sigma(A,B,C)$的能观测指数.

线性系统能观测的判别条件(observability criterion for linear systems) 判别线性系统能观测性的准则. 给定线性时变系统 $\Sigma(A(t),B(t),C(t))$,下述命题等价:

1. $\Sigma(A(t),B(t),C(t))$在 t_0 时刻能观测.

2. 存在某个时刻 $t_1>t_0$,使

$$M(t_1,t_0) = \int_{t_0}^{t_1} \Phi^{\mathrm{T}}(\tau,t_0)C^{\mathrm{T}}(\tau)C(\tau)\Phi(\tau,t_0)\mathrm{d}\tau > 0.$$

3. 存在 $t_1>t_0$,使 $C(t)\Phi(t,t_0)$在$[t_0,t_1]$上列线性独立,其中 $\Phi(t,\tau)$是相应齐次方程的基本解矩阵.

对线性定常系统 $\Sigma(A,B,C)$下述命题等价:

1. $\Sigma(A,B,C)$能观测.

2. 对任给的 $t_1>0$,均有

$$M(t_1,t_0) = \int_{t_0}^{t_1} \mathrm{e}^{A^{\mathrm{T}}(\tau-t_0)}C^{\mathrm{T}}C\mathrm{e}^{A(\tau-t_0)}\mathrm{d}\tau > 0.$$

3. $\mathrm{rank}[C^{\mathrm{T}}\ A^{\mathrm{T}}C^{\mathrm{T}}\ \cdots (A^{\mathrm{T}})^{n-1}C^{\mathrm{T}}]^{\mathrm{T}}=n$.

4. $\mathrm{rank}[sI-A^{\mathrm{T}}\ C^{\mathrm{T}}]^{\mathrm{T}}=n$,
$\forall s\in\sigma(A)=\{s|\det(sI-A)=0\}$.

5. $N\triangleq\bigcap_{k=0}^{n-1}\ker(CA^k)=\{0\}$.

上式中符号 $\ker(\cdot)$为算子的核空间. 判据 5 是系统 $\Sigma(A,B,C)$能观测性的一种几何形式的判据.

对偶原理(duality principle) 一个线性系统的能控性与另一个对应线性系统能观测性间的相互关系. 对给定的线性定常系统 $\Sigma(A,B,C)$,称系统 $\Sigma'(-A^{\mathrm{T}},C^{\mathrm{T}},B^{\mathrm{T}})$为 $\Sigma(A,B,C)$的对偶系统. 显然 $\Sigma(A,B,C)$亦是$\Sigma'(-A^{\mathrm{T}},C^{\mathrm{T}},B^{\mathrm{T}})$的对偶系统. 对偶原理揭示了原系统的能控(或能观测)性和对偶系统

的能观测(或能控)性之间关系. 具体表述为:

1. 系统 $\Sigma(A,B,C)$ 能控的充分必要条件是其对偶系统 $\Sigma'(-A^{\mathrm{T}},C^{\mathrm{T}},B^{\mathrm{T}})$ 为能观测的.

2. 系统 $\Sigma(A,B,C)$ 能观测的充分必要条件是其对偶系统 $\Sigma'(-A^{\mathrm{T}},C^{\mathrm{T}},B^{\mathrm{T}})$ 为能控的.

等价系统(equivalent systems) 两个线性系统在坐标变换下的等价关系. 给定连续线性时变系统 $\Sigma(A(t),B(t),C(t))$, 设 $T(t)$ 是一 n 阶满秩方阵且它的元是 t 的连续可微函数. 在坐标变换 $\hat{x}=T(t)x$ 下, $\Sigma(A(t),B(t),C(t))$ 变为 $\hat{\Sigma}(\hat{A}(t),\hat{B}(t),\hat{C}(t))$, 其中

$$\hat{A}(t)=(T(t)A(t)+\dot{T}(t))T^{-1}(t),$$
$$\hat{B}(t)=T(t)B(t),$$
$$\hat{C}(t)=C(t)T^{-1}(t),$$

此时称 $\hat{\Sigma}(\hat{A}(t),\hat{B}(t),\hat{C}(t))$ 为 $\Sigma(A(t),B(t),C(t))$ 的代数等价系统, 简称等价系统. 记 $\hat{\Phi}(t,t_0)$, $\Phi(t,t_0)$ 为对应 $\hat{A}(t)$, $A(t)$ 的基本解阵, 则

$$\hat{\Phi}(t,t_0)=T(t)\Phi(t,t_0)T^{-1}(t_0).$$

当 $A(t)=A,B(t)=B,C(t)=C,T(t)=T$ 时, 线性定常系统 $\Sigma(A,B,C)$ 的等价系统为 $\hat{\Sigma}(\hat{A},\hat{B},\hat{C})$, 其中

$$\hat{A}=TAT^{-1},\ \hat{B}=TB,\ \hat{C}=CT^{-1}.$$

直接计算可知: $\Sigma(A,B,C)$ 和等价系统 $\hat{\Sigma}(\hat{A},\hat{B},\hat{C})$ 不但有相同的特征多项式、特征方程、极点集、传递函数矩阵和脉冲响应矩阵, 而且其输入解耦零点集、输出解耦零点集和输入-输出解耦零点集都是相同的.

控制系统的标准形(canonical form for control systems) 描述线性系统状态空间的一种典型形式. 状态空间描述形式的线性定常系统

$$\begin{cases}\dot{x}=Ax+Bu,\\ y=Cx,\end{cases}\tag{1}$$

可以通过坐标变换 $x=T\bar{x}$, $|T|\neq 0$, 使系统变成一种新的表达形式

$$\begin{cases}\dot{\bar{x}}=T^{-1}AT\bar{x}+T^{-1}Bu,\\ y=CT\bar{x}.\end{cases}\tag{2}$$

在这一形式下, $T^{-1}AT$, $T^{-1}B$ 或 CT 有特别明显的形式来判定该系统的一些基本性质(如能控性、能观性…). 称(2)式为(1)式的标准形. 除了上述易对系统进行分析的功能外, 还容易在此标准形下对该系统进行设计, 以使闭环系统具有人们所需要的性能.

线性系统的标准形(canonical form for linear systems) 描述线性定常系统能控、能观测性质的状态空间的结构形式. 线性定常系统 $\Sigma(A,B,C)$ 的标准形是从能控、能观测结构出发, 通过选取特殊线性变换而得到的一类代数等价(或相似)系统. 它有能控、能观测标准形和卡尔曼分解等形式. 下面叙述前两种.

如果 $\operatorname{rank}[B\ AB\ \cdots\ A^{n-1}B]=n_1<n$, 则存在非奇异阵 $\widetilde{T}$, 使 $\Sigma(A,B,C)$ 代数等价于

$$\widetilde{\Sigma}:\ \begin{cases}\begin{pmatrix}\dot{\widetilde{x}}_1\\ \dot{\widetilde{x}}_2\end{pmatrix}=\begin{pmatrix}\widetilde{A}_{11} & \widetilde{A}_{12}\\ 0 & \widetilde{A}_{22}\end{pmatrix}\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix}+\begin{pmatrix}\widetilde{B}_1\\ 0\end{pmatrix}u,\\ y=[\widetilde{C}_1\ \ \widetilde{C}_2][\widetilde{x}_1^{\mathrm{T}}\ \ \widetilde{x}_2^{\mathrm{T}}]^{\mathrm{T}},\end{cases}$$

其中 $\widetilde{x}_1\in\mathbb{R}^{n_1}$, $(\widetilde{A}_{11},\widetilde{B}_1)$ 能控. $\widetilde{\Sigma}$ 称为 $\Sigma(A,B,C)$ 的能控标准形.

如果 $\operatorname{rank}[C^{\mathrm{T}}\ A^{\mathrm{T}}C^{\mathrm{T}}\ \cdots\ (A^{\mathrm{T}})^{n-1}C^{\mathrm{T}}]=\hat{n}_1<n$, 则存在非奇异阵, 使 $\Sigma(A,B,C)$ 代数等价于

$$\hat{\Sigma}:\ \begin{cases}\begin{pmatrix}\dot{\hat{x}}_1\\ \dot{\hat{x}}_2\end{pmatrix}=\begin{pmatrix}\hat{A}_{11} & 0\\ \hat{A}_{21} & \hat{A}_{22}\end{pmatrix}\begin{pmatrix}\hat{x}_1\\ \hat{x}_2\end{pmatrix}+\begin{pmatrix}\hat{B}_1\\ \hat{B}_2\end{pmatrix}u,\\ y=\hat{C}_1\hat{x}_1,\end{cases}$$

其中 $\hat{x}_1\in\mathbb{R}^{\hat{n}_1}$, $(\hat{A}_{11},\hat{C}_1)$ 能观测. $\hat{\Sigma}$ 称为 Σ 的能观测标准形.

显然在 $\widetilde{\Sigma}$ 中 $\widetilde{x}_2$ 是不能控的, 而在 $\hat{\Sigma}$ 中 $\hat{x}_2$ 是不能观测的.

单输入系统能控标准形(controllable canonical form for single input systems) 能控单输入系统的一种标准形. 对单输入线性定常系统

$$\Sigma:\ \begin{cases}\dot{x}=Ax+bu,\\ y=Cx,\end{cases}$$

如果它能控, 则存在坐标变换 $x=\overline{T}\bar{x}$, 使

$$\overline{A}=\overline{T}^{-1}A\overline{T}=\begin{pmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & 1\\ -a_n & -a_{n-1} & \cdots & -a_2 & -a_1\end{pmatrix},$$

$$\bar{b}=\overline{T}^{-1}b=\begin{pmatrix}0\\0\\ \vdots\\0\\1\end{pmatrix}.$$

称具有上述形式的系统

$$\begin{cases}\dot{\bar{x}}=\overline{A}\bar{x}+\bar{b}u,\\ y=C\overline{T}\bar{x}\end{cases}$$

为原系统 Σ 的能控相伴标准形, 简称能控标准形.

单输出系统能观标准形(observable canonical form for single output systems) 能观测单输出系统的一种标准形. 对单输出线性定常系统

$$\Sigma:\ \begin{cases}\dot{x}=Ax,\\ y=Cx,\end{cases}$$

如果它能观测, 则容易选取 $\widetilde{T}$, 使在坐标变换 $x=\widetilde{T}\widetilde{x}$ 下有

$$\widetilde{A}=\widetilde{T}^{-1}A\widetilde{T}$$

$$=\begin{bmatrix}0&0&\cdots&0&-\tilde{a}_n\\1&0&\cdots&0&-\tilde{a}_{n-1}\\0&1&\cdots&0&-\tilde{a}_{n-2}\\\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&\cdots&1&-\tilde{a}_1\end{bmatrix},$$

$$\widetilde{C}=C\widetilde{T}=[0\quad 0\quad\cdots\quad 0\quad 1].$$

称具有上述形式的系统

$$\begin{cases}\dot{\tilde{x}}=\widetilde{A}\tilde{x},\\y=\widetilde{C}\tilde{x}\end{cases}$$

为原系统 Σ 的能观相伴标准形,简称能观标准形.

龙伯格能控标准形(Luenberger controllable canonical form) 线性定常系统的一种能控标准形.对能控的线性定常系统

$$\Sigma:\quad\begin{cases}\dot{x}=Ax+Bu,\\y=Cx,\end{cases}$$

可以通过坐标变换变为

$$\begin{cases}\dot{\bar{x}}=\begin{bmatrix}\overline{A}_{11}&\overline{A}_{12}&\cdots&\overline{A}_{1n_1}\\\overline{A}_{21}&\overline{A}_{22}&\cdots&\overline{A}_{2n_1}\\\vdots&\vdots&&\vdots\\\overline{A}_{n_11}&\overline{A}_{n_12}&\cdots&\overline{A}_{n_1n_1}\end{bmatrix}\bar{x}+\begin{bmatrix}\overline{B}_1\\\overline{B}_2\\\vdots\\\overline{B}_{n_1}\end{bmatrix}u,\\y=[\overline{C}_1\quad\overline{C}_2\quad\cdots\quad\overline{C}_{n_1}]\bar{x},\end{cases}$$

这里

$$\overline{A}_{ii}=\begin{bmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&1\\\times&\times&\times&\cdots&\times\end{bmatrix}\in\mathbf{R}^{d_i\times d_i}\quad(\forall i\in\underline{n}_1),$$

$$\overline{A}_{ij}=\begin{bmatrix}0&0&0&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&0\\\times&\times&\times&\cdots&\times\end{bmatrix}\in\mathbf{R}^{d_i\times d_j}$$

$$(\forall i\neq j;\ i,j\in\underline{n}_1),$$

$$\overline{B}_i=\begin{bmatrix}0&0&0&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&0\\0&1&\times&\cdots&\times\end{bmatrix}\in\mathbf{R}^{d_i\times r}\quad(\forall i\in\underline{n}_1).$$

称具有上述形式的系统为原系统 Σ 的多变量的龙伯格能控标准形,其中 n_1 为阵 B 的秩,在将原系统的 B 经过适当的列交换后可使

$$d_1\geqslant d_2\geqslant\cdots\geqslant d_{n_1}(\geqslant 1).$$

龙伯格能控标准形是龙伯格(Luenberger,D. G.)于 1967 年提出的.

龙伯格能观标准形(Luenberger observable canonical form) 线性定常系统的一种能观标准形.对能观测的线性定常系统

$$\Sigma:\quad\begin{cases}\dot{x}=Ax,\\y=Cx,\end{cases}$$

可以通过坐标变换变为

$$\begin{cases}\dot{\tilde{x}}=\begin{bmatrix}\widetilde{A}_{11}&\widetilde{A}_{12}&\cdots&\widetilde{A}_{1\tilde{n}_1}\\\widetilde{A}_{21}&\widetilde{A}_{22}&\cdots&\widetilde{A}_{2\tilde{n}_1}\\\vdots&\vdots&&\vdots\\\widetilde{A}_{\tilde{n}_11}&\widetilde{A}_{\tilde{n}_12}&\cdots&\widetilde{A}_{\tilde{n}_1\tilde{n}_1}\end{bmatrix}\tilde{x},\\y=[\widetilde{C}_1\quad\widetilde{C}_2\quad\cdots\quad\widetilde{C}_{\tilde{n}_1}]\tilde{x},\end{cases}$$

这里

$$\widetilde{A}_{ii}=\begin{bmatrix}0&0&\cdots&0&\times\\1&0&\cdots&0&\times\\0&1&\cdots&0&\times\\\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&\cdots&1&\times\end{bmatrix}\in\mathbf{R}^{\tilde{d}_i\times\tilde{d}_i}\quad(\forall i\in\underline{\tilde{n}}_1),$$

$$\widetilde{A}_{ij}=\begin{bmatrix}0&\cdots&0&\times\\\vdots&&\vdots&\vdots\\0&\cdots&0&\times\end{bmatrix}\in\mathbf{R}^{\tilde{d}_i\times\tilde{d}_j}$$

$$(\forall i\neq j;i,j\in\underline{\tilde{n}}_1),$$

$$\widetilde{C}_i=\begin{bmatrix}0&\cdots&0&0\\0&\cdots&0&1\\0&\cdots&0&\times\\\vdots&&\vdots&\vdots\\0&\cdots&0&\times\end{bmatrix}\in\mathbf{R}^{m\times\tilde{d}_i}\quad(\forall i\in\underline{\tilde{n}}_1),$$

称具有上述形式的系统为原系统 Σ 的多变量的龙伯格能观标准形,$\tilde{n}_1$ 为阵 C 的秩,对 C 阵的行经过适当的行交换后可使

$$\tilde{d}_1\geqslant\tilde{d}_2\geqslant\cdots\geqslant\tilde{d}_{\tilde{n}}(\geqslant 1).$$

龙伯格能观标准形是龙伯格(Luenberger,D. G.)于 1967 年提出的.

线性系统的卡尔曼分解(Kalman decomposition for linear systems) 线性定常系统在相似变换下的一种标准形.给定线性定常系统 $\Sigma(A,B,C)$,其卡尔曼(又称十六块)分解是指它在特殊线性变换下的标准结构型.如果

$$\mathrm{rank}[C^{\mathrm{T}}\quad A^{\mathrm{T}}C^{\mathrm{T}}\quad\cdots\quad(A^{\mathrm{T}})^{n-1}C^{\mathrm{T}}]^{\mathrm{T}}<n,$$

$$\mathrm{rank}[B\quad AB\quad\cdots\quad A^{n-1}B]<n,$$

则一定存在线性变换 T,使 $\Sigma(A,B,C)$ 代数等价(或相似)于

$$\begin{bmatrix}\dot{x}_1\\\dot{x}_2\\\dot{x}_3\\\dot{x}_4\end{bmatrix}=\begin{bmatrix}A_{11}&0&A_{13}&0\\A_{21}&A_{22}&A_{23}&A_{24}\\0&0&A_{33}&0\\0&0&A_{43}&A_{44}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}+\begin{bmatrix}B_1\\B_2\\0\\0\end{bmatrix}u,$$

$$y=[C_1\quad 0\quad C_3\quad 0][x_1^{\mathrm{T}}\quad x_2^{\mathrm{T}}\quad x_3^{\mathrm{T}}\quad x_4^{\mathrm{T}}]^{\mathrm{T}},$$

其中

$$\left(\begin{pmatrix}A_{11}&0\\A_{21}&A_{22}\end{pmatrix}\begin{pmatrix}B_1\\B_2\end{pmatrix}\right)$$

为能控的,而

$$\left(\begin{pmatrix}A_{11}&A_{13}\\0&A_{33}\end{pmatrix}[C_1\,C_3]\right)$$

为能观的.

上式表明：x_1 是既能控又能观的；x_2 是能控不能观的；x_3 是不能控但能观的；x_4 是既不能控又不能观的. 这使人们从结构上对线性系统 $\Sigma(A,B,C)$ 有了更本质的认识.

线性系统的稳定性(stability for linear system) 自由线性系统的一种结构性质. 下列不受控系统

$$\dot{x} = A(t)x \quad (1)$$

称为自由线性系统，其中 $A(t)=[a_{ij}(t)]$ 为 $n\times n$ 矩阵函数，它的元 $a_{ij}(t)(i,j=1,2,\cdots,n)$ 为定义在$[0,+\infty]$上的连续(或分段连续)函数. $x=0$ 是线性系统(1)的平衡状态，平衡状态 $x=0$ 的稳定性、(一致)渐近稳定性、(全局)一致渐近稳定性的定义见"李亚普诺夫稳定性"中自由非线性系统的相应定义. 可以证明：对于系统(1)，上述几种稳定性之间有下列关系：渐近稳定等价于全局渐近稳定；一致渐近稳定等价于全局一致渐近稳定.

当 $A(t)\equiv A$(常数矩阵)时，自由线性定常系统的稳定性之间有下列关系：稳定等价于一致稳定；渐近稳定等价于一致渐近稳定，等价于全局渐近稳定，等价于全局一致渐近稳定.

对于自由线性系统(1)，简称其平衡状态 $x=0$ 的稳定性为系统(1)的稳定性.

自由线性系统稳定性的判别条件(criterion of stability for free linear system) 判别自由线性系统稳定性的准则. 常用的判别条件有：

1. "线性系统的稳定性"条目中线性系统(1)一致渐近稳定的充分必要条件是它按指数稳定，即存在正数 α,β，使得下列不等式成立：

$$\| x(t) \| \leqslant \beta \| x_0 \| \exp[-\alpha(t-t_0)].$$

2. 设 $\lambda_i(t)(i=1,2,\cdots,n)$ 为矩阵 $A(t)+A^{\mathrm{T}}(t)$ 的特征值. 如果 $\mathrm{Re}(\lambda_i(t))\leqslant -h<0$，则"线性系统的稳定性"条目中线性系统(1)是一致渐近稳定的.

当 $A(t)\equiv A$(常数矩阵)时，下列结论成立：

1. 如果 A 的所有特征值皆具有负实部，则"线性系统的稳定性"条目中线性系统(1)渐近稳定.

2. 如果 A 的实部为零的特征值皆是单重的，而其余特征值皆具有负实部，则"线性系统的稳定性"条目中线性系统(1)是临界稳定的.

3. 如果 A 有实部为正的特征值，则"线性系统的稳定性"条目中线性系统(1)不稳定.

4. 如果 A 的实部为零的重特征值的代数重数都等于其几何重数，而其余特征值皆具有负实部，则"线性系统的稳定性"条目中线性系统(1)临界稳定，否则"线性系统的稳定性"(1)不稳定.

贝尔曼-格朗沃尔不等式(Bellman-Gronwall inequality) 用于估计线性积分不等式解的一个不等式. 设函数 $y(t)$满足积分不等式：

$$y(t) \leqslant \int_0^t \alpha(\tau)\beta(\tau)\mathrm{d}\tau + \beta(t) \quad (\forall t \geqslant 0),$$

式中 $\alpha(t),\beta(t)$是已知实函数，则下列贝尔曼-格朗沃尔不等式

$$y(t) \leqslant \int_0^t \alpha(\tau)b(\tau)\exp\left(\int_\tau^t \alpha(s)\mathrm{d}s\right)\mathrm{d}\tau + \beta(t)$$

对$\forall t\geqslant 0$ 成立. 如果 $\beta(t)$是可微的，则有

$$y(t) \leqslant \beta(0)\exp\left(\int_0^t \alpha(\tau)\mathrm{d}\tau\right) + \int_0^t \dot{\beta}(\tau)\exp\left(\int_\tau^t \alpha(s)\mathrm{d}s\right)\mathrm{d}\tau \quad (\forall t \geqslant 0).$$

特别地，当 $\beta(t)$恒为常数时，则有

$$y(t) \leqslant \beta(0)\exp\left(\int_0^t \alpha(\tau)\mathrm{d}\tau\right) \quad (\forall t \geqslant 0).$$

李亚普诺夫方程(Liapunov equation) 一种用于研究线性定常系统李亚普诺夫稳定性的方程. 李亚普诺夫方程为如下的线性矩阵方程：

$$PA + A^{\mathrm{T}}P = -Q,$$

其中 $A,Q=Q^{\mathrm{T}}$ 为已知 $n\times n$ 常数矩阵，P 为未知矩阵，n 为某正整数. 李亚普诺夫方程在线性定常系统

$$\dot{x} = Ax$$

稳定性理论及其应用中起着重要作用，即有：

1. 记矩阵 A 的特征值为 $\lambda_i(i=1,2,\cdots,n)$，任给一对称矩阵 Q，李亚普诺夫方程具有惟一对称解矩阵 P 的充分必要条件是

$$\lambda_i + \lambda_j \neq 0 \ (i,j = 1,2,\cdots,n).$$

2. 对于给定的正定对称矩阵 Q，李亚普诺夫方程具有正定对称解矩阵 P 的充分必要条件是 A 的特征值皆具有负实部.

3. 对于给定的半正定对称矩阵 $Q=C^{\mathrm{T}}C$，当$[A,C^{\mathrm{T}}]$能观测时，李亚普诺夫方程具有正定对称解矩阵 P 的充分必要条件是 A 的特征值皆具有负实部.

4. 李亚普诺夫方程的正定对称解矩阵 P 有如下表达式：

$$P = \int_0^\infty \mathrm{e}^{A^{\mathrm{T}}t}Q\mathrm{e}^{At}\mathrm{d}t.$$

自由线性定常系统的李亚普诺夫函数(Liapunov function for free linear time-invariant system) 用于研究自由线性定常系统稳定性的具有特殊符号性质的标量函数. 对于自由线性定常系统

$$\dot{x} = Ax \quad (1)$$

令 P 为"李亚普诺夫方程"条目中李亚普诺夫方程的正定对称解矩阵，这时二次型标量函数 $v(x)=x^{\mathrm{T}}Px$ 是(1)式的李亚普诺夫函数(参见"李亚普诺夫函数").

周期系数线性系统稳定性判据(criterion of stability for linear system with period coefficient) 判别周期系数线性系统稳定性的准则. 由下列微分

方程描述的系统

$$\dot{x} = A(t)x,\ A(t+T) = A(t) \quad (1)$$

称为周期系数线性系统，其中 $T>0$ 为系统的周期．记 $X(t)$ 为系统(1)的任一基本解矩阵，则易证 $X(t+T)$ 也是系统(1)的一个基本解矩阵，因此存在常值可逆阵 Q，使得 $X(t+T)=X(t)Q$．可以证明：可逆阵 Q 可表示为指数矩阵形式 $Q=e^{A_0T}$，其中 A_0 为常数矩阵．定义 $L(t)=e^{A_0t}X^{-1}(t)$，则通过李亚普诺夫变换

$$y = L(t)x \quad (2)$$

可将系统(1)化为常系数线性系统

$$\dot{y} = A_0 y. \quad (3)$$

周期系数线性系统(1)的稳定性判据为：

1．如果 A_0 的所有特征值皆具有负实部，则系统(1)是渐近稳定的．

2．如果 A_0 至少有一个实部为正的特征值，则系统(1)不稳定．

3．如果 A_0 的实部为零的特征值都是单重的，则系统(1)是临界稳定的．

4．如果 A_0 的实部为零的特征值的代数重数都等于其几何重数而其余特征值皆具有负实部，则系统(1)是临界稳定的，否则是不稳定的．

极点(poles)　反映线性定常系统稳定性的基本概念．是线性定常系统特征多项式的零点．给定线性定常系统 $\Sigma(A,B,C)$，称 $\det(sI-A)$ 为系统 Σ 的特征多项式，而称 $\det(sI-A)=0$ 为系统 Σ 的特征方程．特征方程的根，在矩阵论中称为 A 的特征值，而在线性系统理论中称为系统 Σ 的极点．由于特征多项式的系数皆为实数，因此如果系统 Σ 在极点中存在复数必共轭成双．通常记

$$\sigma(A) = \{s \mid \det(sI - A) = 0\}$$

为系统 Σ 的极点集合．系统 Σ 的极点决定了系统 Σ 的部分动态性能，如稳定性、衰减特性、稳定裕度等．

极点配置(pole placement)　控制系统设计的重要概念．系统 $\Sigma(A,B,C)$ 的极点决定系统的部分动态性能．然而当系统 $\Sigma(A,B,C)$ 给定后，其极点分布不一定满足设计者的要求，通过状态反馈来移动 Σ 的极点使其满足设计要求，就是控制系统 Σ 的极点配置问题．对给定的线性定常系统 $\Sigma(A,B,C)$ 和任意给定的由 n 个复数组成的集合 Λ（其中复数必共轭成双），如果存在状态反馈使得闭环系统极点

$$\sigma(A + BK) = \Lambda,$$

则称系统 $\Sigma(A,B,C)$ 是可以任意极点配置的．系统 $\Sigma(A,B,C)$ 可以任意极点配置的充分必要条件是系统 $\Sigma(A,B,C)$ 为完全能控的．

状态反馈(state feedback)　控制系统设计中的一种基本控制方式．给定线性定常系统 $\Sigma(A,B,C)$，称

$$u = Kx + Fv \quad (1)$$

为 Σ 的状态反馈，其中 $K\in \mathbb{R}^{r\times n}$，$F\in \mathbb{R}^{r\times r}$，且 $\det F\neq 0$．将(1)式代入原系统得闭环系统 $\Sigma_C(A+BK,BF,C)$．Σ_C 表明状态反馈改变了原系统 Σ 的结构，因此也就改变了原系统的动态行为．状态反馈不能改变原系统 Σ 的能控性，但能改变原系统 Σ 的能观测性．

对系统 $\Sigma(A,B,C)$，量测的通常不是全部状态 x 而是输出 y．此时为实现状态反馈必须增加状态观测器．为了直接应用量测信息，取 $u=Hy+Fv$，其中 $H\in \mathbb{R}^{r\times m}$，$F$ 同上，通常称为输出反馈．输出反馈既不改变原系统 Σ 的能控性，又不改变原系统 Σ 的能观测性．

能稳性(stabilizability)　控制系统的结构性质之一．对给定的线性定常系统 $\Sigma(A,B,C)$，如果存在状态反馈 $u = Kx$，其中 $K \in \mathbb{R}^{r\times n}$，使得闭环系统是渐近稳定的，即 $\sigma(A + BK) \subset \mathbb{C}^-$，则称系统 $\Sigma(A,B,C)$ 是能稳的．有时也称矩阵对(A,B) 是能稳的．系统 $\Sigma(A,B,C)$ 能稳的充分必要条件是系统 $\Sigma(A,B,C)$ 的不能控极点都是稳定的．系统 $\Sigma(A,B,C)$ 的能稳性概念在系统镇定问题的讨论中有重要作用．

系统的镇定(stabilization of system)　控制系统设计中的一个概念．指设计控制器使得闭环系统稳定．具有这类性质的控制器中最简单、最常用的一种是状态反馈，即对给定的系统 $\Sigma(A,B,C)$，选择状态反馈形式的控制器 $u=Kx$，$K\in \mathbb{R}^{r\times n}$，使得 $\sigma(A+BK)\triangleq\{s\mid\det(sI-A-BK)=0\}\subset\mathbb{C}^-$，$\mathbb{C}^-$ 为左半开复平面．寻找满足上述条件的状态反馈增益阵 K 的问题称为系统 $\Sigma(A,B,C)$ 的镇定．虽然系统 Σ 的镇定和系统 Σ 的任意极点配置都是求反馈增益阵 K，但前者要求选取 K 使得闭环系统的极点位于左半开复平面内；而后者则要求选择 K 使得闭环系统的极点位于复平面内的任意给定位置上．因此，它们是控制系统设计中的两个不同概念，且前者比后者的能解性条件弱．系统 $\Sigma(A,B,C)$ 能镇定的充分必要条件是系统 $\Sigma(A,B,C)$ 为能稳的（参见“能稳性”）．

状态重构(state reconstruction)　从量测信息中获得状态的过程．对给定的线性定常系统 $\Sigma(A,B,C)$，通常能为控制器提供的可用信号是输出 y 而不是状态 x．然而有些控制系统设计方法中所需要的控制形式又恰是状态反馈 $u=Kx$．因此，为了实现状态反馈，人们必须寻找一种方法，使得状态 x 能从量测 $y=Cx$ 所提供的信息中构造出来，这就是状态重构问题．设 $\Sigma(A,B,C)$ 是能观测的，即可观测矩阵 R_0 满足 $\mathrm{rank}\,R_0=n$（n 为系统阶数），直接计算可得

$$x=(R_0R_0^{\mathrm{T}})^{-1}\sum_{i=0}^{n-1}(A^i)^{\mathrm{T}}C^{\mathrm{T}}[y^{(i)}-CB\,u^{(i-1)}-\cdots-CA^{i-1}Bu],$$

其中$u^{(i-1)}=0$(当$i-1<0$时). 上式表明状态x可用输出y控制u及其导数$y^{(i)}$,$u^{(i)}$的线性组合来表达. 但由于量测y中总存在噪声,而微分对噪声又有放大作用,致使从理论上合理的上述状态重构方法,在实际上是不可取的. 尽管如此,它却为人们构造具有动态性能的状态观测器提供了有力的启示.

状态观测器(state observer) 由系统的输出信息获取状态估计的算法或装置. 由"状态重构"条目中知道,即使系统$\Sigma(A,B,C)$是能观测的,要确切获得其状态x也是很困难的. 因此只能退而求状态x的一个渐近估计,即通过一个动态系统的动态过程来逐渐获得状态的全部信息. 该动态系统以原系统的输出$y(t)$和输入$u(t)$作为输入,而其输出便是状态x的一种渐近估计. 这个动态系统就称为状态观测器(亦称状态估计器或状态重构器),最先由龙伯格(Luenberger,D. G.)提出. 基本形式为

$$\dot{x}_{\mathrm{C}}=(A-GC)x_{\mathrm{C}}+Bu+Gy, \qquad (1)$$

其中$G\in\mathbf{R}^{n\times m}$是待定的,只要取$G$使$\sigma(A-GC)\subset\mathbf{C}^-$($\mathbf{C}^-$为左半开复平面),便有

$$\lim_{t\to\infty}(x(t)-x_{\mathrm{C}}(t))=0,$$

即$x_{\mathrm{C}}(t)$是$x(t)$的渐近估计. 由于$x_{\mathrm{C}}\in\mathbf{R}^n$,(1)式称为全阶状态观测器. 系统$\Sigma(A,B,C)$存在状态观测器的充分必要条件为$\Sigma(A,B,C)$是能检测的.

最小阶状态观测器(state observer with minimal order) 阶数最低的状态观测器. 给定线性定常系统$\Sigma(A,B,C)$,当它的量测输出y能提供一部分状态信息时,可以为其设计低阶的状态观测器. 在系统$\Sigma(A,B,C)$的所有状态观测器中阶数最低者称为最小阶状态观测器. 如果$\Sigma(A,B,C)$能观测且

$$\operatorname{rank}C=\operatorname{rank}[C_1,C_2]=m,$$

其中C_1是满秩阵,则存在$n-m$阶状态观测器,它是最小阶的. 具体形式为

$$\dot{z}=(A_{22}-G_2A_{12})z+(B_2-G_2B_1)u+[(A_{21}-G_2A_{11})+(A_{22}-G_2A_{12})G_2]y,$$

而

$$x_{\mathrm{C}}=\begin{pmatrix}C_1^{-1}(I-C_2G_2) & -C_1^{-1}C_2\\ G_2 & I\end{pmatrix}\begin{pmatrix}y\\ z\end{pmatrix},$$

其中

$$\begin{pmatrix}C_1 & C_2\\ 0 & I\end{pmatrix}A\begin{pmatrix}C_1^{-1} & -C_1^{-1}C_2\\ 0 & I\end{pmatrix}=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},$$

$$\begin{pmatrix}C_1 & C_2\\ 0 & I\end{pmatrix}B=\begin{pmatrix}B_1\\ B_2\end{pmatrix},$$

$G_2\in\mathbf{R}^{(n-m)\times m}$是使$\sigma(A_{22}-G_2A_{12})\subset\mathbf{C}^-$的待定矩阵,$\mathbf{C}^-$表示左半开复平面.

分离性原理(separation principle) 状态反馈和状态观测器能分开进行设计的原理. 给定既能控又能观的线性定常系统$\Sigma(A,B,C)$,利用其状态观测器

$$\dot{x}_{\mathrm{C}}=(A-GC)x_{\mathrm{C}}+Bu+Gy$$

的x_{C}代替状态反馈控制$u=Kx$中的x,以构成动态补偿器

$$\Sigma_{\mathrm{p}}:\ \begin{cases}\dot{x}_{\mathrm{C}}=(A-GC)x_{\mathrm{C}}+Bu+Gy,\\ u=Kx_{\mathrm{C}}.\end{cases}$$

联合Σ_{p}和$\Sigma(A,B,C)$得Σ_{C}:

$$\begin{pmatrix}\dot{x}\\ \dot{x}_{\mathrm{C}}\end{pmatrix}=\begin{pmatrix}A & BK\\ GC & A-GC+BK\end{pmatrix}\begin{pmatrix}x\\ x_{\mathrm{C}}\end{pmatrix}.$$

直接计算得

$$\det\begin{pmatrix}sI-A & -BK\\ -GC & sI-A+GC-BK\end{pmatrix}=\det(sI-A-BK)\det(sI-A+GC),$$

即

$$\sigma\begin{pmatrix}A & BK\\ GC & A-GC+BK\end{pmatrix}=\sigma(A+BK)\cup\sigma(A-GC).$$

通常称上述关系为分离性原理,即状态观测器的加入对状态反馈的作用毫无影响. 因此可独立设计状态反馈(选K)和状态观测器(选G).

线性系统的动态补偿(dynamic compensation for linear system) 一种控制器的设计方法. 它是基于系统量测输出构造另一线性动态系统并以其输出来设计系统的反馈控制器的过程. 亦称为动态反馈过程. 线性控制系统的设计是根据量测(输出)来选择控制器以改造被控对象,使其具有人们所希望的品质. 控制器按结构分为两类:一类是静态输出反馈;另一类是动态输出反馈. 习惯上把为系统设计静态反馈控制器的过程称为反馈,而把为系统设计动态反馈控制器的过程称为动态补偿. 通常静态反馈简单,但它对被控对象的改造能力往往不能令人满意;而动态补偿虽然复杂,但它却能使被改造后的系统具有优良的品质. 具体设计时,应根据对系统的品质要求、能提供的设备和经费综合考虑,以确定控制器的类型.

动态补偿器(dynamic compensator) 亦称动态输出反馈控制器. 一种控制器形式. 它是动态补偿过程中设计出的系统反馈控制器. 其形式为

$$\Sigma_{\mathrm{p}}:\ \begin{cases}\dot{x}_{\mathrm{C}}=A_{\mathrm{C}}x_{\mathrm{C}}+B_{\mathrm{C}}y,\\ u=F_{\mathrm{C}}x_{\mathrm{C}}+Fy,\end{cases}$$

其中$x_{\mathrm{C}}\in\mathbf{R}^{n_{\mathrm{C}}}$,$A_{\mathrm{C}}$,$B_{\mathrm{C}}$,$F_{\mathrm{C}}$,$F$都是适当维数的常数矩阵. 所谓动态补偿器$\Sigma_{\mathrm{p}}$的设计,就是根据对闭环系统的品质要求来选择$A_{\mathrm{C}}$,$B_{\mathrm{C}}$,$F_{\mathrm{C}}$,$F$四个常数矩阵. 显然在能保证闭环系统品质要求的条件下,应使n_{C}尽量小,以降低控制器造价和增强系统的可靠性. 在

“分离性原理”条目中，利用状态反馈和状态观测器构成的控制器就是一个动态补偿器. 据 n_C 尽量小的要求，无疑应选择最小阶状态观测器.

动态输出反馈控制器(dynamic output feedback compensator) 即“动态补偿器”.

线性系统解耦问题(decoupling problem for linear system) 线性控制系统理论的一个基本问题. 对线性定常系统

$$\Sigma:\ \begin{cases}\dot{x} = Ax + Bu,\ x \in \mathbb{R}^n,\ u \in \mathbb{R}^r,\\ y = Cx, \qquad\qquad y \in \mathbb{R}^p,\end{cases}$$

如果存在两矩阵 K 及 H，使得在控制 $u=Kx+Hv$ 作用下，闭环系统

$$\begin{cases}\dot{x} = (A + BK)x + BHv,\\ y = Cx\end{cases}$$

的传递函数阵为非奇异对角方阵，则称系统 Σ 的解耦问题有解(简称能解耦). 而(K,H)称为该问题的解. 此问题是由摩根(Morgan,B. S.)于 1964 年提出的，因此也称为摩根问题. 该问题至今尚未彻底解决.

对 $r=p$(方系统)的情形，问题有解的充分必要条件为

$$D = \begin{pmatrix} c_1A^{l_1-1}B\\ \vdots \\ c_pA^{l_p-1}B\end{pmatrix}$$

非奇异，这里 c_i 为 C 的第 i 行 n 维行向量，

$$l_i = \min\{j|c_iA^{j-1}B \neq 0\},\forall i \in \{1,2,\cdots,p\}.$$

对 $r>p$ 的情形，到目前(1999 年初)为止，该问题有解的充分必要条件仍未得到.

线性系统块解耦问题(block decoupling problem for linear system) 一类控制设计问题. 它是通过状态反馈及输入变换，使闭环系统的输入、输出间形成一“组”与一“组”的对应关系的问题. 对线性定常系统

$$\begin{cases}\dot{x}=Ax+Bu,\\ y=Cx,\end{cases}$$

若存在阵 K 与 $G(|G|\neq 0)$，使 $C(sI-A-BK)^{-1}BG$ 是块对角阵

$$\text{block diag}\{W_1(s),W_2(s),\cdots,W_l(s)\},$$

这里 $W_i(s)$是 $p_i\times r_i$ 阶非零有理分式阵($p_i\geqslant 1,r_i\geqslant 1$)且 $l\geqslant 2$，则称系统对分割 $P=\{p_1,p_2,\cdots,p_l\}$的块解耦问题有解. 系统(A,B,C)的能控部分必可严格等价于多项式阵描述形式的能控标准形$(P(s),B_1,R(s))$. 其中

$$\begin{bmatrix}P(s)\\ R(s)\end{bmatrix}$$

是列次为$\{d_1,d_2,\cdots d_r\}$的列首一多项式阵. 引入

$$R(s) = \begin{pmatrix}R_1(s)\\ R_2(s)\\ \vdots \\ R_l(s)\end{pmatrix},$$

$R_i(s)$为 p_i 行多项式阵.

$$M_i(s) = R_i(s)\cdot \text{diag}\{s^{d_1-d_1},s^{d_1-d_2},\cdots,s^{d_1-d_r}\},$$

$$U_i(s)M_i(s) = \begin{bmatrix}\overline{M}_i(s)\\ 0\end{bmatrix},\quad \overline{M}_i(s)\ \text{行正则},$$

$U_i(s)$是 $p_i\times p_i$ 阶单位模阵. $\overline{E}_i=\Gamma_r(\overline{M}_i(s))$，则系统$(A,B,C)$对分割 $P=\{p_1,p_2,\cdots,p_l\}$的块解耦问题有解的充分必要条件是

$$\overline{E} = \begin{pmatrix}\overline{E}_1\\ \overline{E}_2\\ \vdots\\ \overline{E}_l\end{pmatrix}$$

行满秩.

注：多项式阵 $M(s)$的第 j 列的列次是指 $M(s)$的第 j 列诸元素中次数最大者，即

$$\alpha_j = \partial_{c_j}(M(s)) = \max_i(\partial(m_{ij}(s))),$$

这时，$M(s)$可表示成

$$M(s)=M_0\begin{pmatrix}s^{\alpha_1} & 0 & \cdots & 0\\ 0 & s^{\alpha_2} & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & s^{\alpha_l}\end{pmatrix}
+M_1\begin{pmatrix}s^{\alpha_1-1} & 0 & \cdots & 0\\ 0 & s^{\alpha_2-1} & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & s^{\alpha_l-1}\end{pmatrix}
+\cdots+M_\alpha\begin{pmatrix}s^{\alpha_1-\alpha} & 0 & \cdots & 0\\ 0 & s^{\alpha_2-\alpha} & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & s^{\alpha_l-\alpha}\end{pmatrix},$$

这里 $\alpha=\max\{\alpha_1,\alpha_2,\cdots,\alpha_l\}$. 而称 M_0 为 $M(s)$的列次系数阵，记为 $\Gamma_C(M(s))=M_0$. 当

$$\partial_{c_j}(M(s)) \geqslant \partial_{c_{j+1}}(M(s)),\ \forall j \in \{1,\cdots,l-1\},$$

且

$$\Gamma_C(M(s)) = \begin{bmatrix}I_l\\ 0\end{bmatrix}$$

时，称 $M(s)$为列首一多项式阵.

状态空间形式线性系统的抗干扰性(disturbance resistance of linear system in state space form) 现代控制理论中所研究的一种控制系统特性. 它是反映系统输出不受外干扰影响能力的性质. 控制系统中总存在外干扰，通常分为有模型和无模型两类. 称系统的输出不受外干扰影响的能力为系统的抗干扰性. 给定线性定常系统

$$\Sigma:\quad \begin{cases} \dot{x}_1 = A_1x_1 + A_3x_2, \\ y = C_1x_1 + C_2x_2 \end{cases}$$

和外干扰 x_2 满足的方程

$$\Sigma_d:\quad \dot{x}_2 = A_2x_2,$$

其中 $x_1 \in \mathrm{R}^{n_1}$ 是状态，$\sigma(A_2) \subset \mathrm{C}^+$，$\mathrm{C}^+$ 表示右半闭复平面. 如果对 Σ_d 的任一解 $x_2(t)$，系统 Σ 都存在一个解 $x_1(t)$，使得

$$y(t) = C_1x_1(t) + C_2x_2(t) = 0,$$

则称系统 Σ 是能抗 Σ_d 型外干扰的. 系统 Σ 能抗 Σ_d 型外干扰的充分必要条件是存在 $L \in \mathrm{R}^{n_1 \times n_2}$，使：

1. $C_2 = -C_1L$.

2. 矩阵 $A_1L - LA_2 + A_3$ 的每一列都是 Σ 的不能观测状态.

结构稳定性(structural stability)　一种稳定性质. 它是系统的结构参数发生扰动时系统仍能保持稳定的性质. 给定系统

$$\Sigma:\quad \dot{x} = Ax.$$

设 $\sigma(A) \subset \mathrm{C}^-$，$\mathrm{C}^-$ 表示左半开复平面. 将矩阵 A 的元素按一定次序排列成一个向量，记为 $p(A) \in \mathrm{R}^{n^2}$，$p(A)$ 称为系统 Σ 的标称参数点. 如果在空间 R^{n^2} 中存在 $p(A)$ 的一个邻域 $U_\delta(p(A))$，只要结构参数 A 扰动后对应的向量 $p(A+\Delta A) \in U_\delta(p(A))$，皆有 $\sigma(A+\Delta A) \subset \mathrm{C}^-$，则称系统 Σ 关于结构参数 A 在 $p(A)$ 处是结构稳定的.

鲁棒性(robustness)　现代控制系统的一个重要概念. 它是刻画系统参数或结构变化以及受外干扰作用时，系统性质是否保持的一个概念. 设系统 Σ 具有某种性质 π. 当其结构参数发生扰动后，系统 Σ 的性质 π 仍能保持，则称系统 Σ 的性质 π 具有鲁棒性. 严格讲，设系统 Σ 具有性质 π. 记 $p(\Sigma) \in \mathrm{R}^l$ 为系统 Σ 的标称参数点. 如果在 R^l 中 $p(A)$ 处存在一个邻域 $U_\delta(p(\Sigma))$. 只要系统 Σ 结构参数扰动后对应的参数向量属于 $U_\delta(p(\Sigma))$ 时，扰动后的系统皆具有性质 π，则称系统 Σ 的性质 π 具有鲁棒性.

稳健性(robustness)　即“鲁棒性”.

线性系统的干扰解耦(disturbance decoupling of linear system)　一种控制设计问题. 它是设计状态反馈使外干扰对系统输出不产生影响. 给定带外干扰 x_2 的系统

$$\Sigma:\quad \begin{cases} \dot{x}_1 = A_1x_1 + B_1u + A_3x_2, \\ y = C_1x_1, \end{cases}$$

$x_2(t) \in \mathscr{X}_2 \triangleq \{x_2(t) \mid x_2(t)$ 是定义在$[0,\infty)$上的连续函数$\}$，如果存在状态反馈 $u = Kx_1$，使得闭环系统对给定的 $x_1(0) \in \mathrm{R}^{n_1}$ 和任意的 $x_2 \in \mathscr{X}_2$ 都有相同的输出 $y(t)$，即外干扰不影响输出，则称输出 y 关于 x_2 是解耦的. 如果记 $\mathscr{K} = \ker C_1$，$\varphi = \mathrm{Im}\, A_3$，则 y 关于 x_2 解耦的充分必要条件为 $\langle A_1 + B_1K \mid \varphi \rangle \subset \mathscr{K}$. 记包含在 $\mathscr{K}$ 中的最大 (A_1, B_1) 不变子空间为 $\mathscr{V}^*$，则 y 关于 x_2 解耦问题有解的充分必要条件是 $\mathscr{V}^* \supset \varphi$.

注：子空间 $\mathscr{S}$ 称为是 (A,B) 不变的系指 $A(\mathscr{S}) \subset \mathscr{S} + \mathrm{Im}\, B$，而 $\langle A \mid \mathscr{B} \rangle = \mathscr{B} + A\mathscr{B} + \cdots + A^{n-1}\mathscr{B}$.

线性调节器(linear regulator)　一种控制器形式. 它是能使闭环系统渐近稳定的线性状态反馈控制器. 给定线性定常系统 $\Sigma(A,B)$，可以通过系统镇定方法产生线性调节器，亦可以求解如下最优控制问题来获得：

$$\begin{cases} \dot{x} = Ax + Bu, \\ J(u(\cdot)) = \int_0^\infty [x^{\mathrm{T}}Qx + u^{\mathrm{T}}Ru]\mathrm{d}t = \min, \end{cases}$$

这里 $Q \geqslant 0$，$R > 0$. 其最优控制为 $u^* = Kx$，$K = -R^{-1}B^{\mathrm{T}}P$. 而 P 满足代数黎卡提方程

$$PA + A^{\mathrm{T}}P + Q - PBR^{-1}B^{\mathrm{T}}P = 0.$$

设 $Q = C^{\mathrm{T}}C$ 是 Q 的任一分解. 代数黎卡提方程存在惟一非负定解 P，使得 $\sigma(A - BR^{-1}B^{\mathrm{T}}P) \subset \mathrm{C}^-$ 的充分必要条件是 (A,B) 能稳、(A,C) 能检测. $\mathbf{C}^-$ 表示左半开复平面.

线性系统的输出调节(output regulation of linear system)　一种控制的设计方法. 它是利用输出信号与动态补偿信号实现输出稳定的方法. 在调节(或跟踪)系统设计中，由于对被调节量 z 要求它具有渐近性质，因此，时时刻刻消除外干扰对被调节量的影响实无必要，只要随着调节进程渐近地消除外干扰对被调节量的影响就够了. 这正是输出调节所要讨论的问题. 给定系统

$$\Sigma:\quad \begin{cases} \dot{x}_1 = A_1x_1 + B_1u + A_3x_2, \\ \dot{x}_2 = A_2x_2, \\ y = C_1x_1 + C_2x_2, \\ z = D_1x_1 + D_2x_2, \end{cases}$$

其中 $\sigma(A_2) \subset \mathrm{C}^+$，$\mathrm{C}^+$ 表示右半闭复平面. 系统 Σ 的动态补偿器 Σ_p 为

$$\Sigma_p:\quad \begin{cases} \dot{x}_C = A_Cx_C + B_Cy, \\ u = F_Cx_C + Fy, \end{cases}$$

联合 Σ_p 和 Σ 得 Σ_C. 当 $x_2 = 0$ 时，如果 Σ_C 是稳定的，则称 Σ_C 是内部稳定的；如果 Σ_C 是内部稳定的，且对任意的 $x_2(0) \in \mathrm{R}^{n_2}$ 皆有

$$\lim_{t\to\infty} z(t) = 0,$$

则称 Σ_C 是输出调节的. 此时称 Σ_p 为系统 Σ 的稳定综合.

线性系统的内模原理(internal model principle for linear system)　设计线性控制系统的一种原理. 它是为受外干扰系统设计的使闭环具有结构稳定抗干扰能力的控制器的综合准则. 对带外干扰的调节(或跟踪)系统 Σ(参见“线性系统的输出调节”)，为保证闭环系统 Σ_C 具有结构稳定的抗外干扰

能力,需要为其设计结构稳定综合 Σ_p. 内模原理提供了设计结构稳定综合 Σ_p 的准则:即在 Σ_p 中必含外干扰的动力学模型. 具体叙述为:Σ_p 是系统 Σ 在数据点 $p\{A_1,B_1,C_1,F_C,B_C,F,A_3,C_2,D_1,D_2\}$处的结构稳定综合的充分必要条件为:

1. (A_1,B_1)能稳,(A_1,C_1)能检测.

2. z 能从 y 中读出.

3. Σ_C 是内部稳定的.

4. Σ_p 中编入了 A_2 的一个内模,这个内模关于 z 能控,关于 u 能观测.

注:z 能从 y 中读出可理解为 z 是 y 的一部分. Σ_p 中编入 A_2 的内模是指:如果$(s-\lambda_i)^{k_i}$是 A_2 的初等因子,则 A_C 关于$(s-\lambda_i)$的初等因子的次数不低于 k_i.

传递函数矩阵(transfer function matrix) 线性定常系统的一种频域描述方法. 给定线性定常系统 $\Sigma(A,B,C)$. 设 $x(0)=0$,如果记状态 $x(t)$、控制 $u(t)$ 和输出 $y(t)$的拉普拉斯变换分别为 $X(s)$,$U(s)$ 和 $Y(s)$,对状态方程和输出方程两边进行拉氏变换后得到

$$Y(s)=C(sI-A)^{-1}BU(s),$$

其中 s 是拉普拉斯算子. 通常称

$$G(s)\triangleq C(sI-A)^{-1}B$$

为系统 Σ 的 传递函数矩阵. $G(s)$是以 s 的有理分式为元素的 $m\times r$ 矩阵. 传递函数矩阵 $G(s)$是线性定常控制系统的另外一种描述方法. 它是用频域方法分析和设计控制系统的基础.

状态空间的实现(realization of state space) 线性系统频域描述到时域描述的转换. 它是将传递函数矩阵用状态空间形式描述出来. 其中,阶数最低的状态空间描述称为最小实现. 给定线性定常系统 $\Sigma(A,B,C)$,易得其传递函数矩阵. 反过来,任给一严格真有理分式矩阵 $G(s)\in\mathrm{R}^{m\times r}(s)$,如果存在 $A\in\mathrm{R}^{n\times n}$,$B\in\mathrm{R}^{n\times r}$,$C\in\mathrm{R}^{m\times n}$,使得

$$G(s)=C(sI-A)^{-1}B,$$

则称 n 阶系统 $\Sigma(A,B,C)$ 为 $G(s)$ 的一个状态空间实现,简称实现. 在 $G(s)$的实现中阶数最低者称为 $G(s)$的最小实现. 已知 $\Sigma(A,B,C)$ 是最小实现的充分必要条件为(A,B)能控,(A,C)能观测. 利用严格真有理分式和其相伴形之关系可直接得其单输入-单输出实现. 将此用于给定的严格真有理分式矩阵 $G(s)\in\mathrm{R}^{m\times r}(s)$的每个元素,可得 mr 个单输入-单输出实现,然后将这 mr 个单输入-单输出实现进行适当组合便得 $G(s)$的实现. 从 $G(s)$的任一实现中找出其既能控又能观测的子系统,便是 $G(s)$的最小实现. 当然由 $G(s)$的汉克尔矩阵,按确定步骤亦可获得 $G(s)$的最小实现.

状态空间的最小实现(minimum realization of state space) 见"状态空间的实现".

系统零点(zero of system) 一种描述控制理论中输入输出的基本概念. 它是系统各种类型零点的总称. 给定线性定常系统 $\Sigma(A,B,C)$,称

$$S(s)=\begin{pmatrix} sI-A & -B \\ C & 0 \end{pmatrix}$$

为 $\Sigma(A,B,C)$ 的系统矩阵. 设

$$\operatorname{rank}S(s)=n+q,\quad q\leqslant\min(m,r),$$

则形如

$$S(s)\begin{pmatrix} i_1,i_2,\cdots,i_q \\ j_1,j_2,\cdots,j_q \end{pmatrix}$$

(它是由 $S(s)$中的第 $1,2,\cdots,n,n+i_1,n+i_2,\cdots,n+i_q$ 行和第 $1,2,\cdots,n,n+j_1,n+j_2,\cdots,n+j_q$ 列元素组成矩阵的行列式)的所有子式之最大公因式的零点称为系统零点,其全体记为 Z_S. Z_S 具有如下性质:

$$Z_S=Z_G\cup Z_d=Z_G\cup\{Z_{i,d}\cup Z_{o,d}\backslash Z_{i,o,d}\},$$

其中 Z_G 是传递函数矩阵的零点集;$Z_{i,d}$ 是输入解耦零点集;$Z_{o,d}$ 是输出解耦零点集;$Z_{i,o,d}$ 是输入-输出解耦零点集.

输入解耦零点(input decouple zero) 一类系统零点. 它是反映系统不能控部分的零点. 对线性定常系统 $\Sigma(A,B,C)$,称使

$$\operatorname{rank}[\lambda I-A,B]<n$$

的复数 λ 为系统 Σ 的输入解耦零点,其全体记为 $Z_{i,d}$,它包含于 A 的特征值集 $\sigma(A)$. 系统的输入解耦零点一定是系统的不可控极点,因此,系统 Σ(完全)能控的充分必要条件是该系统没有输入解耦零点.

输出解耦零点(output decouple zero) 一类系统零点. 它是反映系统不能观测部分的零点. 对线性定常系统 $\Sigma(A,B,C)$,称使

$$\operatorname{rank}\begin{bmatrix} \lambda I-A \\ C \end{bmatrix}<n$$

的复数 λ 为系统 Σ 的输出解耦零点,其全体记为 $Z_{o,d}$,它包含于 A 的特征值集 $\sigma(A)$. 系统的输出解耦零点一定是系统的不可观测极点,因此,系统 Σ(完全)能观测的充分必要条件是该系统不存在输出解耦零点.

输入-输出解耦零点(input-output decouple zero) 一类系统零点. 它是反映系统不能控和不能观部分的零点. 给定线性定常系统 $\Sigma(A,B,C)$,称同时满足如下条件:

$$\operatorname{rank}[\lambda I-A,B]<n,$$

$$\operatorname{rank}\begin{bmatrix} \lambda I-A \\ C \end{bmatrix}<n$$

的复数 λ 为系统 $\Sigma(A,B,C)$的输入-输出解耦零点,其全体记为 $Z_{i,o,d}$. 每一个输入-输出零点都既是系统

$\Sigma(A,B,C)$ 的不能控极点，又是它的不能观测极点．

传递函数矩阵的零点(zero of transfer function matrix)　一类系统零点．它是反映系统的既能控、又能观测部分的零点．任给有理分式矩阵 $G(s)\in \mathrm{R}^{m\times r}(s)$，总存在左、右互质分解

$$G(s)=P_l^{-1}(s)N_l(s)=N_r(s)P_r^{-1}(s),$$

其中 $P_l(s)(P_r(s))$和 $N_l(s)(N_r(s))$是左(右)互质的多项式矩阵．$P_l(s)$，$P_r(s)$称为 $G(s)$的分母，$N_l(s)$，$N_r(s)$称为 $G(s)$的分子．

设 $\mathrm{rank}\,G(s)=k\leqslant\min(m,r)$，$N(s)$为 $G(s)$的任一分子，且其史密斯型为

$$\begin{pmatrix}\varepsilon_1(s) & 0 & \cdots & 0 & 0\\ 0 & \varepsilon_2(s) & \cdots & 0 & 0\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & \varepsilon_k(s) & 0\\ 0 & 0 & \cdots & 0 & 0\end{pmatrix}\in \mathrm{R}^{m\times r}[s].$$

称

$$Z_G\triangleq\{s\,|\,\varepsilon_i(s)=0,\ i=1,2,\cdots,k\}$$

为 $G(s)$的零点集合，其中的元素简称 $G(s)$的零点．

由于 $N_l(s)$和 $N_r(s)$有相同的史密斯型，因此 $G(s)$的零点可惟一确定．

系统的传输零点(transmission zero of system)　系统的一类零点．给定线性定常系统 $\Sigma(A,B,C)$，如果

$$\mathrm{rank}\,S(s)=\mathrm{rank}\begin{pmatrix}sI-A & -B\\ C & 0\end{pmatrix}<n+\min(m,r),$$

则称系统 $\Sigma(A,B,C)$为退化的；否则称为非退化的．使 $\mathrm{rank}S(\bar{s})<n+\min(m,r)$的复数 $\bar{s}$ 称为系统 $\Sigma(A,B,C)$的传输零点，其全体记为 Z_T．当系统 $\Sigma(A,B,C)$退化时，$Z_T=\mathrm{C}$，C 表示整个复平面；当系统 $\Sigma(A,B,C)$是一个既能控、又能观测的非退化系统时，则有

$$Z_T=Z_G=Z_S,$$

其中 Z_G 是系统 $\Sigma(A,B,C)$的传递函数矩阵零点集，而 Z_S 是其系统零点集(参见“系统零点”)．

零极相消(zero-pole cancellation)　系统存在不能控或不能观测部分时的一种现象．给定线性定常系统 $\Sigma(A,B,C)$，除非系统 $\Sigma(A,B,C)$ 是既能控又能观测，一般地，$sI-A$ 和 B 总存在左公因式矩阵，而 $sI-A$ 和 C 总存在右公因式矩阵．在形成传递函数矩阵 $G(s)=C(sI-A)^{-1}B$ 的过程中，必将 $sI-A$ 和 B 的左公因式矩阵及 $sI-A$ 和 C 的右公因式矩阵消去．这种现象称系统 $\Sigma(A,B,C)$的零极相消．消去的零(或极)点集合为 $\sigma(A)\cap(Z_{i.d}\cup Z_{o.d})$，其中 $\sigma(A)=\{s\,|\det(sI-A)=0\}$，$Z_{i.d}$ 是输入解耦零点集；$Z_{o.d}$ 是输出解耦零点集．

线性系统的多项式阵描述形式(polynomial matrix description for linear system)　线性控制系统的一种数学表述形式．用 m 维变量 z 的高阶线性微分方程组

$$P(\mathrm{D})z=Q(\mathrm{D})u$$

叠加上量测方程

$$y=R(\mathrm{D})z+E(\mathrm{D})u$$

组成的系统称为线性系统的多项式阵描述形式，这里$\mathrm{D}(\equiv \mathrm{d}/\mathrm{d}t)$为微分算符，$u$ 及 y 是 r 维控制变量及 p 维量测变量，$P(s)$是 $m\times m$ 阶实系数非奇异多项式方阵，$Q(s)$，$R(s)$与 $E(s)$分别是 $m\times r$ 阶、$p\times m$ 阶与 $p\times r$ 阶实系数多项式阵，它们的元素都是 s 的实多项式，它们的系数最一般的情形可以是时间(自变量)t 的实值函数．当 $P(s)$，$Q(s)$，$R(s)$与 $E(s)$的元素都是 s 的实系数(时不变)多项式时，则上述系统是线性定常系统．在这里，一般称 z 为分状态变量．这样，状态空间的描述形式

$$\begin{cases}\dot{x}=Ax+Bu,\\ y=Cx,\end{cases}$$

也可以作为它的一种特殊形式

$$\begin{cases}(DI-A)x=Bu,\\ y=Cx\end{cases}$$

出现，称

$$\begin{bmatrix}P(s) & -Q(s)\\ R(s) & E(s)\end{bmatrix}$$

为相应系统的系统阵．

多项式阵描述形式系统的等价变换(equivalent transformation for system in polynomial matrix description)　用多项式阵形式描述的同一系统的不同表现形式．严格等价的定义是由罗森布罗克(Rosenbrock，H. H.)给出的．称线性定常系统

$$\begin{cases}P(\mathrm{D})z=Q(\mathrm{D})u,\\ y=R(\mathrm{D})z+E(\mathrm{D})u\end{cases}$$

与系统

$$\begin{cases}\overline{P}(\mathrm{D})\bar{z}=\overline{Q}(\mathrm{D})u,\\ y=\overline{R}(\mathrm{D})\bar{z}+\overline{E}(\mathrm{D})u\end{cases}$$

是严格等价的，如果存在着单位模阵 $U(s)$与 $V(s)$，以及适当阶次的多项式阵 $M(s)$与 $N(s)$，使上两系统经扩充后的系统阵间满足

$$\begin{bmatrix}\begin{bmatrix}\overline{P}(s) & 0\\ 0 & I_1\end{bmatrix} & -\begin{bmatrix}\overline{Q}(s)\\ 0\end{bmatrix}\\ [\overline{R}(s)\quad 0] & \overline{E}(s)\end{bmatrix}$$

$$=\begin{bmatrix}U(s) & 0\\ M(s) & I_p\end{bmatrix}\begin{bmatrix}\begin{bmatrix}I_2 & 0\\ 0 & P(s)\end{bmatrix} & -\begin{bmatrix}0\\ Q(s)\end{bmatrix}\\ [0\quad R(s)] & E(s)\end{bmatrix}$$

$$\cdot\begin{bmatrix}V(s) & N(s)\\ 0 & I_p\end{bmatrix},$$

这里 I_1 与 I_2 和 I_p 为适当行(列)数的单位阵．

注：一个非奇异多项式阵称为单位模阵是指它的行列式是个非零实数.

多项式阵描述形式系统的能控性判据(controllability criterion for system in polynomial matrix description)　一类判断线性定常系统能控性的准则. 线性定常系统

$$P(D)z = Q(D)u$$

完全能控的充分必要条件是：$P(s)$与$Q(s)$左互质，即它们的最大左公因子是单位模阵. 判断两多项式阵$M_1(s)$与$M_2(s)$左互质的等价判据有下列四条：

1. $[M_1(s)\ M_2(s)]$的史密斯标准形为$[I\ 0]$.

2. $[M_1(\lambda)\ M_2(\lambda)]$行满秩，$\forall\lambda\in\mathbb{C}$.

3. 存在适当阶数的多项式阵$X_1(s)$与$X_2(s)$，使

$$[M_1(s)\ M_2(s)]\begin{bmatrix}X_1(s)\\X_2(s)\end{bmatrix}=I.$$

4. 存在适当阶数的多项式阵$Y_1(s)$与$Y_2(s)$，使

$$\begin{bmatrix}M_1(s) & M_2(s)\\Y_1(s) & Y_2(s)\end{bmatrix}$$

为单位模阵.

多项式阵描述形式系统的能观性判据(observability criterion for system in polynomial matrix description)　一类判断线性定常系统能观性的准则. 线性定常系统

$$\begin{cases}P(D)z = Q(D)u,\\y = R(D)z + E(D)u\end{cases}$$

完全能观的充分必要条件是：$P(s)$与$R(s)$右互质，即它们的最大右公因子是单位模阵. 判断两多项式阵$M_1(s)$与$M_2(s)$右互质的等价判据有下列四条：

1. $\begin{bmatrix}M_1(s)\\M_2(s)\end{bmatrix}$的史密斯标准形为$\begin{bmatrix}I\\0\end{bmatrix}$.

2. $\begin{bmatrix}M_1(\lambda)\\M_2(\lambda)\end{bmatrix}$列满秩，$\forall\lambda\in\mathbb{C}$.

3. 存在两个相应阶数的多项式阵$X_1(s)$与$X_2(s)$，使

$$[X_1(s)\ X_2(s)]\begin{bmatrix}M_1(s)\\M_2(s)\end{bmatrix}=I.$$

4. 存在两个相应阶数的多项式阵$Y_1(s)$与$Y_2(s)$，使

$$\begin{bmatrix}M_1(s) & Y_1(s)\\M_2(s) & Y_2(s)\end{bmatrix}$$

为单位模阵.

多项式阵描述形式系统的能控标准形(controllable canonical form for system in polynomial matrix description)　线性定常系统的一种频域标准形. 它是多项式阵描述形式的线性定常系统完全能控时均可等价变换成的形式. 线性定常系统

$$\begin{cases}P(D)z = Q(D)u,\\y = R(D)z + E(D)u\end{cases}$$

完全能控时，必严格等价于下述形式的能控标准形

$$\begin{cases}\bar{P}(D)\bar{z} = \bar{B}_1u,\\y = \bar{R}(D)\bar{z} + \bar{E}(D)u,\end{cases}$$

其中$\bar{B}_1$是行满秩常数矩阵，

$$\begin{bmatrix}\bar{P}(s)\\\bar{R}(s)\end{bmatrix}$$

为列次不小于1的列首一多项式阵.

多项式阵描述形式系统的能观标准形(observable canonical form for system in polynomial matrix description)　线性定常系统的一种频域标准形. 它是多项式阵描述形式的线性定常系统完全能观时均可等价变换成的形式. 线性定常系统

$$\begin{cases}P(D)z = Q(D)u,\\y = R(D)z + E(D)u\end{cases}$$

完全能观时，必严格等价于下述形式的能观标准形

$$\begin{cases}\tilde{P}(D)\tilde{z} = \tilde{Q}(D)u,\\y = \tilde{C}_1\tilde{z} + \tilde{E}(D)u,\end{cases}$$

其中，$\tilde{C}_1$是列满秩常数矩阵，$[\tilde{P}(s)\ \tilde{Q}(s)]$为行次均不小于1的行首一多项式阵.

注：多项式阵$M(s)$为行首一多项式阵是指：$M^T(s)$是列首一多项式阵(列次、列次系数阵、列首一多项式阵的意义见“多项式阵描述形式系统的能控标准形”).

多项式阵描述形式系统的传递函数阵(transfer function matrix for system in polynomial matrix description)　线性定常系统的一种频域表示. 它是描述系统输入、输出间频域特征性关系的一个有理分式阵. 线性定常系统

$$\begin{cases}P(D)z = Q(D)u,\\y = R(D)z + E(D)u\end{cases}$$

的传递函数阵为下述有理分式阵：

$$W(s) = R(s)P^{-1}(s)Q(s) + E(s),$$

它表征输入变量$u(t)$与输出变量$y(t)$的(零初始状态下)拉普拉斯变换$U(s)$和$Y(s)$之间的传递关系，即$Y(s)=W(s)U(s)$.

任一有理分式阵等价于史密斯-麦克米兰标准形(参见“史密斯-麦克米兰形”)，即存在相应阶数的单位模阵$U(s)$与$V(s)$，使

$$U(s)W(s)V(s)=\begin{pmatrix}\frac{\varepsilon_1(s)}{d_1(s)} & 0 & \cdots & 0 & 0\\0 & \frac{\varepsilon_2(s)}{d_2(s)} & \cdots & 0 & 0\\\vdots & \vdots & & \vdots & \vdots\\0 & 0 & \cdots & \frac{\varepsilon_l(s)}{d_l(s)} & 0\\0 & 0 & \cdots & 0 & 0\end{pmatrix},$$

这里，$\varepsilon_i(s)/d_i(s)$是不可约分式，且$\varepsilon_i(s)$是$\varepsilon_{i+1}(s)$的因式，$d_{i+1}(s)$是$d_i(s)$的因式.

多项式阵标准形与状态空间标准形间的等价(equivalence between polynomial matrix canonical form and state space canonical form) 两类系统描述形式间严格等价的具体形式. 线性定常系统

$$\begin{cases} P(\mathrm{D})z = B_1 u, \\ y = R(\mathrm{D})z, \end{cases}$$

如果 B_1 行满秩,

$$\begin{bmatrix} P(s) \\ R(s) \end{bmatrix}$$

为列次不小于 1 的列首一多项式阵,

$$\begin{bmatrix} P(s) \\ R(s) \end{bmatrix} = \begin{bmatrix} I_1 \\ 0 \end{bmatrix} \begin{bmatrix} s^{d_1} & 0 & \cdots & 0 \\ 0 & s^{d_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & s^{d_m} \end{bmatrix} + \begin{bmatrix} A_1 \\ C_1 \end{bmatrix} \begin{bmatrix} s^{d_1-1} & 0 & \cdots & 0 \\ 0 & s^{d_2-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & s^{d_{n_1}-1} \end{bmatrix} + \begin{bmatrix} A_2 \\ C_2 \end{bmatrix} \begin{bmatrix} s^{d_1-2} & 0 & \cdots & 0 \\ 0 & s^{d_2-2} & \cdots & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & s^{d_{n_2}-2} \end{bmatrix} + \cdots + \begin{bmatrix} A_\nu \\ C_\nu \end{bmatrix} [I_\nu \quad 0],$$

则上述完全能控系统严格等价于

$$\begin{cases} \dot{x} = \begin{bmatrix} 0 & (I_\nu\ 0) & 0 & \cdots & 0 \\ 0 & 0 & (I_{\nu-1}\ 0) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & (I_2\ 0) \\ -A_\nu & -A_{\nu-1} & -A_{\nu-2} & \cdots & -A_1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ B_1 \end{bmatrix} u, \\ y = [C_\nu \quad C_{\nu-1} \quad \cdots \quad C_2 \quad C_1] x, \end{cases}$$

这里 I_i 为 n_i 阶单位阵,n_i 为集合

$$\{d_1 - i + 1, d_2 - i + 1, \cdots, d_m - i + 1\}$$

中正整数的个数,满足

$$(m =) n_1 \geqslant n_2 \geqslant \cdots \geqslant n_\nu (\geqslant 1),$$

$$\nu = d_1, \quad \sum_{i=1}^{\nu} n_i = \sum_{j=1}^{m} d_j.$$

类似地,线性定常系统

$$\begin{cases} \overline{P}(\mathrm{D})\bar{z} = \overline{Q}(\mathrm{D})u, \\ y = \overline{C}_1 \bar{z} \end{cases}$$

中,如果 $\overline{C}_1$ 列满秩,$[\overline{P}(s)\,\overline{Q}(s)]$ 为行次不小于 1 的行首一多项式阵,

$$[\overline{P}(s)\ \overline{Q}(s)] = \begin{bmatrix} s^{\alpha_1} & 0 & \cdots & 0 \\ 0 & s^{\alpha_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & s^{\alpha_m} \end{bmatrix} [I_1\ 0] + \begin{bmatrix} s^{\alpha_1-1} & 0 & \cdots & 0 \\ 0 & s^{\alpha_2-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & s^{\alpha_{\bar{n}_1}-1} \end{bmatrix} [A_1\ B_1] + \begin{bmatrix} s^{\alpha_1-2} & 0 & \cdots & 0 \\ 0 & s^{\alpha_2-2} & \cdots & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & s^{\alpha_{\bar{n}_2}-2} \end{bmatrix} [A_2\ B_2] + \cdots + \begin{bmatrix} I_\mu \\ 0 \end{bmatrix} [\overline{A}_\mu\ \overline{B}_\mu],$$

则上述完全能观系统严格等价于

$$\begin{cases} \dot{\bar{x}} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\overline{A}_\mu \\ \begin{pmatrix} I_\mu \\ 0 \end{pmatrix} & 0 & \cdots & 0 & -\overline{A}_{\mu-1} \\ 0 & \begin{pmatrix} I_{\mu-1} \\ 0 \end{pmatrix} & \cdots & 0 & -\overline{A}_{\mu-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \begin{pmatrix} I_2 \\ 0 \end{pmatrix} & -\overline{A}_1 \end{bmatrix} \bar{x} + \begin{bmatrix} \overline{B}_\mu \\ \overline{B}_{\mu-1} \\ \vdots \\ \overline{B}_2 \\ \overline{B}_1 \end{bmatrix} u, \\ y = [0 \quad 0 \quad \cdots \quad 0 \quad \overline{C}_1] \bar{x}, \end{cases}$$

这里 I_j 为 $\bar{n}_j$ 阶单位阵,$\bar{n}_j$ 是集$\{\alpha_1 - j + 1, \alpha_2 - j + 1, \cdots, \alpha_m - j + 1\}$中正整数的个数,满足

$$(m =) \bar{n}_1 \geqslant \bar{n}_2 \geqslant \cdots \geqslant \bar{n}_\mu (\geqslant 1),$$

$$\mu = \alpha_1, \quad \sum_{j=1}^{\mu} \bar{n}_j = \sum_{i=1}^{m} \alpha_i.$$

传递函数阵的实现(realization for transfer function matrix) 描述线性定常系统的一种转换方式. 给定一个有理分式矩阵,求以此矩阵为传递函数阵的对应系统的问题. 给定有理分式阵 $W(s)$ 的实现是指:构造一个系统(状态空间描述形式或多项式阵描述形式),使该系统的传递函数阵为 $W(s)$,则称该系统是 $W(s)$ 的一个实现. 在 $W(s)$ 的所有实现中,阶数最小者被称为是 $W(s)$ 的最小实现. 给定 $W(s)$,求其实现与最小实现的一种方法如下:首先把 $W(s)$ 分解成严格真有理分式 $\overline{W}(s)$ 与多项式阵 $E(s)$ 之和,前者是指其每一个元素的分子多项式次数小于分母多项式次数,这样

$$W(s) = \overline{W}(s) + E(s).$$

再对 $\overline{W}(s)$进行左分解 $\overline{W}(s)=P_1^{-1}(s)Q_1(s)$或右分解 $\overline{W}(s)=\overline{R}_1(s)\overline{P}_1^{-1}(s)$，则

$$\begin{cases}P_1(\mathrm{D})z_1=Q_1(\mathrm{D})u,\\ y=z_1+E(\mathrm{D})u\end{cases}$$

或

$$\begin{cases}\overline{P}(\mathrm{D})\bar{z}_1=u,\\ y=\overline{R}_1(\mathrm{D})\bar{z}_1+E(\mathrm{D})u\end{cases}$$

均是 $W(s)$的实现. 如果其左(右)分解还是左(右)既约分解，则对应的系统就是 $W(s)$的最小实现. 把上述系统化成能观标准形及能控标准形，就可得到相应的状态空间描述形式的实现或最小实现了.

传递函数阵的最小实现(minimum realization for transfer function matrix) 见"传递函数阵的实现".

多项式阵描述形式的系统抗干扰性(disturbance resistance of system in polynomial matrix description) 描述系统对外干扰的抵抗能力. 对 q 维满足方程 $P_0(\mathrm{D})f=0$ 的干扰 f，系统

$$\begin{cases}P(\mathrm{D})z=F(\mathrm{D})f,\\ y=R(\mathrm{D})z+E(\mathrm{D})f\end{cases}$$

的输出 y 恒为零(不受干扰 f 的影响)，则称上述系统能抗 $P_0(s)$型干扰 f.

对给定的 $P(s)$，$F(s)$，$R(s)$，$E(s)$及 $P_0(s)$，系统能抗 $P_0(s)$型干扰的充分必要条件是：存在多项式阵 $L(s)$，使

$$\begin{pmatrix}\overline{P}(s)L(s)-F(s)\\ \overline{R}(s)L(s)+E(s)\end{pmatrix}$$

具有右因子 $P_0(s)$，这里 $\overline{P}(s)$与 $\overline{R}(s)$是 $P(s)$与$R(s)$的右互质部分.

动态补偿器设计的多项式阵法(polynomial matrix method for designing dynamic compensator) 设计系统动态补偿器的一种方法. 即把各种类型的动态补偿器的设计方法归结为解单边丢番图方程的问题. 常用的方法是：

1. 对系统

$$\begin{cases}P(\mathrm{D})z=u,\\ y=R(\mathrm{D})z\end{cases}$$

设计动态补偿器

$$\begin{cases}X(\mathrm{D})z_c=Y(\mathrm{D})y,\\ y_c=z_c\end{cases}$$

及控制律 $u=v-y_c$，则由解下列单边丢番图方程

$$X(s)P(s)+Y(s)R(s)=H(s)$$

而得解 $X(s)$，$Y(s)$，这里 $H(s)$由对闭环系统所需的要求确定.

2. 对系统

$$\begin{cases}\overline{P}(\mathrm{D})\bar{z}=\overline{Q}(\mathrm{D})u,\\ y=\bar{z}\end{cases}$$

设计动态补偿器

$$\begin{cases}\overline{X}_1(\mathrm{D})\bar{z}_{c_1}=y,\\ \bar{y}_{c_1}=\overline{Y}_1(\mathrm{D})\bar{z}_{c_1}\end{cases}$$

及控制律 $u=v-\bar{y}_{c_1}$或设计动态补偿器

$$\begin{cases}\overline{X}_2(\mathrm{D})\bar{z}_{c_2}=e,\\ \bar{y}_{c_2}=\overline{Y}_2(\mathrm{D})\bar{z}_{c_2}\end{cases}$$

及控制律 $u=\bar{y}_{c_2}$，这里 $e=v_0-y$，v_0 为给定的被跟踪信号，则由解下列单边丢番图方程

$$\overline{P}(s)\overline{X}_i(s)+\overline{Q}(s)\overline{Y}_i(s)=\overline{H}_i(s)\quad(i=1,2)$$

而得解 $\overline{X}_i(s)$，$\overline{Y}_i(s)$，这里 $\overline{H}_i(s)(i=1,2)$由相应问题的闭环系统的要求而定.

最小阶动态补偿器(minimal order dynamic compensator) 满足闭环系统特性要求的阶数最小的动态补偿器. 动态补偿器的阶数由 $\deg(\det X(s))$或$\deg(\det \overline{X}_i(s))(i=1,2)$确定. 为使动态补偿器能物理实现，闭环系统的阶数 $\deg(\det H(s))$或 $\deg(\det \overline{H}_i(s))$有一个下界的限制. "动态补偿器设计的多项式阵法"条目中第 1 类动态补偿器的最小阶数为$r(\mu-1)$，这里 r 是控制输入变量的维数，μ 是系统的能观指数；而第 2 类动态补偿器的最小阶数为$p(\nu-1+\delta)$，这里 p 为输出变量的维数，ν 为系统的能控指数，δ 为所需的抗干扰、跟踪等模型的最小公倍式的次数. 对最简单的不具备抗干扰及跟踪能力的动态补偿器，其最小阶动态补偿器的阶数为 $\min\{r(\mu-1),p(\nu-1)\}$.

特征函数(eigenfunction) 矩阵特征值概念的一种推广. 它是以函数为元素的矩阵的特征多项式的根. 设 $G(s)$为函数域上的 $m\times m$ 矩阵，s 为复数自变量，则满足下列方程的函数 $g(s)$即为矩阵 $G(s)$的特征函数

$$\det[g(s)I_m-G(s)]=0.$$

展开上式可得关于 $g(s)$的 m 次代数方程. 特别地，设 $G(s)$的各元素均为 s 的实系数有理函数(例如，$G(s)$为控制系统的传递函数矩阵的情形)，则上式可化为

$$g^m(s)+b_{m-1}(s)g^{m-1}(s)+\cdots+b_1(s)g(s)+b_0(s)=0,$$

式中诸 $b_i(s)$均为 s 的多项式. 如果此方程左端为不可约，则由此方程所确定的特征函数 $g(s)$是 s 的 m 值代数函数. 特征函数 $g(s)$的极点和零点都是传递函数矩阵的史密斯-麦克米兰极点和零点. 但 $G(s)$可能还有一些史密斯-麦克米兰极点和与之重合的零点不是 $g(s)$的极点和零点. 特征函数是研究和设计多变量控制系统的一种重要数学工具.

特征轨迹(eigenlocus) 用特征函数在复平面上刻画系统频域特性的曲线. 它是 $m\times m$ 矩阵 $G(s)$的特征函数 $g(s)$顺时针绕闭域 $\Omega\triangleq\{\lambda|\lambda\in\mathrm{C},\mathrm{Re}\,\lambda\geqslant$

0}的周界 Γ 一周时画出的轨迹(参见"特征函数"). 如果 $G(s)$ 在虚轴上有极点,则上述闭合曲线应沿无穷小的半圆从右侧绕过这些极点. 一般情形下特征轨迹为 m 支闭合的曲线. 根据控制系统开环传递函数矩阵的特征轨迹包围复数平面上 $-1+\mathrm{j}$ 点的周数,可以判断闭环系统的稳定性(参见"多变量系统的奈奎斯特稳定性判据"). 当 $m=1$,特征轨迹就退化为标量函数 $G(s)$ 的奈奎斯特图象.

多变量系统的奈奎斯特稳定性判据(Nyquist stability criterion for multivariable system) 亦称广义奈奎斯特稳定性判据. 一种判别多输入-多输出系统稳定性的频域准则. 它是单变量系统的奈奎斯特稳定性判据的重要推广,它根据多变量系统的开环传递函数矩阵判断闭环系统是否稳定. 设 $m\times m$ 的开环传递函数矩阵 $G(s)$ 在右半开复数平面上有 p 个极点($p\geqslant 0$),则闭环系统稳定的充分必要条件为 $G(s)$ 的特征轨迹反时针包围 $-1+\mathrm{j}$ 点 p 周. 当 $m=1$,此判据即退化为单变量系统的奈奎斯特稳定性判据.

广义奈奎斯特稳定性判据(generalized Nyquist stability criterion) 即"多变量系统的奈奎斯特稳定性判据".

奇异值(singular value) 与矩阵相联系的一组非负实数. 对于任一 $m\times n$ 复矩阵 G,存在 $m\times m$ 酉矩阵 U 以及 $n\times n$ 酉矩阵 V,使得

$$G = UDV^*, \tag{1}$$

这里上标"$*$"表示共轭转置,D 为 $m\times n$ 实对角矩阵并具有形式

$$D=\begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix},$$

其中 $D_r=\mathrm{diag}\ (\sigma_1,\sigma_2,\cdots,\sigma_r)$,且 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0$. 称矩阵 D 的对角线上的元素为矩阵 G 的奇异值;称 $\sigma_1,\sigma_2,\cdots,\sigma_r$ 为矩阵 G 的正奇异值. 通常,记 $\bar{\sigma}=\sigma_1$,$\underline{\sigma}=\sigma_r$.

称(1)式为 G 的奇异值分解. 一个矩阵的奇异值是惟一确定的,但矩阵 U,V 则不惟一. 当 G 是实矩阵时,可选 U,V 为实的正交矩阵. 利用奇异值分解可以定义矩阵的条件数 $\mathrm{cond}(G)=\bar{\sigma}/\underline{\sigma}\geqslant 1$,可以构造矩阵的一种"伪逆". 另外,

$$\|Gx\|_2/\|x\|_2\leqslant\bar{\sigma},$$

这里 x 为任意 n 维向量. 由此可见,$\bar{\sigma}$ 给出了矩阵 G 作用于各向量上所可能产生的范数放大倍数的上界.

奇异值分解对于鲁棒控制系统的分析与设计有重要作用.

正规矩阵(normal matrix) 与自身的共轭转置矩阵可交换的方阵. 矩阵为正规的充分必要条件是其特征向量矩阵为酉矩阵. 由于摄动下方矩阵的特征值的偏移量的上限正比于矩阵的特征向量矩阵的条件数,而酉矩阵的条件数最小(等于 1),故正规矩阵对于设计鲁棒控制系统有重要用途. 控制系统的某些传递函数矩阵若为正规矩阵,则系统的 H_∞ 范数意义下的鲁棒性达到最佳.

并矢矩阵(dyadic matrix) 非奇异多项式矩阵的一种特殊表示形式. 当

$$G(s)=\sum_{i=1}^{m}g_i(s)l_ir_i^{\mathrm{T}},$$

其中函数 $g_i(s)\neq 0$,l_i,r_i 为相应维数的列向量,称 $G(s)$ 为并矢矩阵. 上式称为 $G(s)$ 的并矢展开. 可知

$$G(s)=L\mathrm{diag}\{g_1(s),g_2(s),\cdots,g_m(s)\}R,$$

这里 $L=[l_1,l_2,\cdots,l_m]$,$R=[r_1,r_2,\cdots,r_m]^{\mathrm{T}}$.

一般地,如果有

$$G^{-1}(s)=A_0s+A_1$$

或

$$G^{-1}(s)=s(A_0s+A_1),$$

其中 A_0 和 A_1 为实常数矩阵,A_0 非奇异,则 $G(s)$ 为并矢矩阵. 如果被控制对象的传递函数矩阵是并矢矩阵,则可用并矢展开方法为它设计控制系统.

结构奇异值(structured singular value) 鲁棒控制理论中的一个概念. 它是与矩阵的最大奇异值有关的一个正数. 亦称为 μ 值. 其定义如下:设 M 为 $n\times n$ 的复数矩阵,令

$$X_\delta=\{\mathrm{block\ diag}(\overbrace{\Delta_1,\cdots,\Delta_1}^{m_1},\overbrace{\Delta_2,\cdots,\Delta_2}^{m_2},\cdots,\overbrace{\Delta_r,\cdots,\Delta_r}^{m_r})\mid\Delta_i\ 为\ k_i\times k_i\ 方阵,\ \bar{\sigma}(\Delta_i)<\delta,\ i=1,2,\cdots,r\},$$

其中 r,m_i,k_i 为正整数,且 $\sum_{i=1}^{r}m_ik_i=n$,用以表示摄动的结构,$\delta>0$,$\bar{\sigma}(\cdot)$ 表示矩阵的最大奇异值. 再令

$$X_\infty=\bigcup_{j=1}^{\infty}X_j,$$

它表示所有最大奇异值没有限制的按上述结构的块对角矩阵的集合,则矩阵 M 的结构奇异值为

$$\mu(M)=\begin{cases}0, & 若不存在\ \Delta\in X_\infty,使\ \det(I+M\Delta)=0;\\ \left(\min\limits_{\Delta\in X_\infty}\{\bar{\sigma}(\Delta)\mid\det(I+M\Delta)=0\}\right)^{-1}, & 其他.\end{cases}$$

当 $r=1,m_1=1$ 时,$\mu(M)=\bar{\sigma}(M)$.

结构奇异值对研究控制系统在结构明确的摄动下的鲁棒性问题有重要作用(参见"μ 理论").

史密斯形(Smith form) 多项式矩阵的一种规范形式. 常用于分析多项式矩阵的零点,求两个多项式矩阵的最大公因子,判断两个多项式矩阵的互质性,求有理分式矩阵的史密斯-麦克米兰形等. 任一多项式矩阵通过初等变换都可化为史密斯形. 其定义如下:设有 $m\times n$ 的多项式矩阵 $N(s)$,其秩为

$r\leqslant\min(m,n)$，必可用 $m\times m$ 的单位模矩阵 $L(s)$ 和 $n\times n$ 的单位模矩阵 $R(s)$，把 $N(s)$ 变成其史密斯形 $S(s)$：

$$S(s)=L(s)N(s)R(s)$$

$$=\left(\begin{array}{cccc|c} S_1(s) & 0 & \cdots & 0 & \\ 0 & S_2(s) & \cdots & 0 & \\ \vdots & \vdots & & 0 & 0_{r\times(n-r)} \\ 0 & 0 & \cdots & S_r(s) & \\ \hline & 0_{(m-r)\times r} & & & 0_{(m-r)\times(n-r)} \end{array}\right),$$

式中诸 $S_i(s)$ $(i=1,2,\cdots,r)$ 为首一多项式，称为 $N(s)$ 的不变因子，且具有依次可整除性质，即 $S_i(s)|S_{i+1}(s)$，$i=1,\cdots,r-1$.

史密斯-麦克米兰形(Smith-McMillan form) 有理分式矩阵的一种规范形式. 可用于研究有理分式矩阵的零点和极点等性质，特别是它为分析多变量系统的传递零点和极点提供了理论工具. 其定义如下：设 $G(s)$ 为 $m\times n$ 的有理分式矩阵，其秩为

$$r\leqslant\min(m,n).$$

$$M(s)=L(s)G(s)R(s)$$

$$=\left(\begin{array}{cccc|c} \frac{\varepsilon_1(s)}{\psi_1(s)} & 0 & \cdots & 0 & \\ 0 & \frac{\varepsilon_2(s)}{\psi_2(s)} & \cdots & 0 & \\ \vdots & \vdots & & \vdots & 0_{r\times(n-r)} \\ 0 & 0 & \cdots & \frac{\varepsilon_r(s)}{\psi_r(s)} & \\ \hline & 0_{(m-r)\times r} & & & 0_{(m-r)\times(n-r)} \end{array}\right),$$

式中 $L(s)$，$R(s)$ 为单位模矩阵，$\psi_1(s)$ 是 $G(s)$ 诸元素分母的最小公倍式，每一对 $\varepsilon_i(s)$ 与 $\psi_i(s)$ 为互质的首一多项式，并有依次可整除性：

$$\varepsilon_i(s)|\varepsilon_{i+1}(s),\ \psi_{i+1}(s)|\psi_i(s)$$
$$(i=1,2,\cdots,r-1).$$

$G(s)$ 的零点就是诸 $\varepsilon_i(s)$ 的零点的全体，$G(s)$ 的极点就是诸 $\psi_i(s)$ 零点的全体.

最小阶系统(least order system) 对应于给定传递函数矩阵阶次最低的系统. 此时，描述线性定常系统的实现与最小实现相同. 系统的微分方程组和代数方程组的一般的形式为

$$\begin{cases}T(s)\bar{\xi}=U(s)\bar{u},\\ \bar{y}=V(s)\bar{\xi}+W(s)\bar{u},\end{cases}$$

式中 $\bar{u}$，$\bar{y}$，$\bar{\xi}$ 分别为向量 u，y，ξ 的拉普拉斯变换，u 为 l 维输入向量，y 为 m 维输出向量，ξ 为 r 维状态向量. T，U，V，W 为 s 的多项式矩阵，维数分别为

$$r\times r,\ r\times l,\ m\times r,\ m\times l.$$

上式可写为

$$\begin{pmatrix}T(s) & U(s)\\ -V(s) & W(s)\end{pmatrix}\begin{pmatrix}\bar{\xi}\\ -\bar{u}\end{pmatrix}=\begin{pmatrix}0\\ -\bar{y}\end{pmatrix},$$

式中

$$\begin{pmatrix}T(s) & U(s)\\ -V(s) & W(s)\end{pmatrix}\triangleq P(s)$$

称为罗森布罗克系统矩阵. 对于给定的传递函数矩阵 $G(s)$，满足

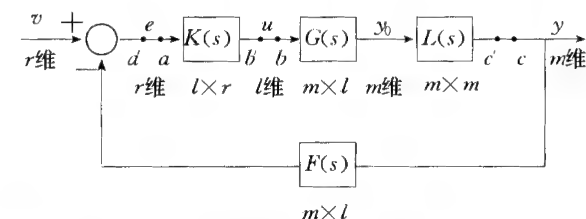
$$G(s)=V(s)T^{-1}(s)U(s)+W(s)$$

的系统矩阵是非惟一的，每一个都称为 $G(s)$ 的一个系统矩阵实现. 当

$$T(s)=sI-A,\quad U(s)=B,$$
$$V(s)=C,\quad W(s)=D(s)$$

时，$P(s)$ 称为状态空间系统矩阵. 在 $G(s)$ 的所有实现中，阶最低的系统矩阵都是严格等价的，称为给定的 $G(s)$ 的最小阶系统. 它们满足 $T(s)$，$U(s)$ 为左互质，即没有输入解耦零点；又满足 $T(s)$，$V(s)$ 为右互质，即没有输出解耦零点，因而是状态完全可控和完全可观测的，且结构最简单的系统.

回比矩阵(return-ratio matrix) 分析多变量控制系统稳定性能和动态性能的一种重要工具. 在图示的多变量控制系统中，v 为输入向量，e 为误差向量，u 为控制向量，y_0 为对象输出向量，y 为系统输出向量，$G(s)$ 为被控制对象的传递函数矩阵，$K(s)$ 和 $L(s)$ 分别为控制器和输出补偿器的传递函



数矩阵，$F(s)$ 为传感器的传递函数矩阵. 若在 a' 到 a 之间断开闭环，并在 a 点加入信号，而在 a' 点取出信号，则从 a 点沿回路到 a' 点的传递函数矩阵反号后定义为该系统关于 a 点的回比矩阵，记为

$$H_{oe}(s)=F(s)L(s)G(s)K(s).$$

回比矩阵相当于单变量控制系统的开环传递函数. 但回比矩阵(及其维数)因断点位置不同而不同. 若在 b' 到 b 之间断开闭环，则仿上可定义该系统关于 b 点的回比矩阵为

$$H_{ou}(s)=K(s)F(s)L(s)G(s).$$

同理，系统关于 c 点的回比矩阵为

$$H_{oy}(s)=L(s)G(s)K(s)F(s).$$

回差矩阵(return-difference matrix) 分析多变量控制系统的一种数学工具(参见"回比矩阵"及其附图). 若在"回比矩阵"条目的图中 a' 到 a 点之间断开闭环,并在 a 点加入信号 e_a,再从 a' 点取出信号 $e_{a'}$,则此两信号之差为

$$\bar{e}_a - \bar{e}_{a'} = [I_r + F(s)L(s)G(s)K(s)]\bar{e}_a = [I_r + H_{oe}(s)]\bar{e}_a,$$

式中 $\bar{e}_a$ 和 $\bar{e}_{a'}$ 分别表示 e_a 和 $e_{a'}$ 的拉普拉斯变换. $H_{oe}(s)$ 为关于 a 点的回比矩阵. 关于 a 点的回差矩阵即定义为

$$D_e(s) \triangleq [I_r + H_{oe}(s)].$$

同理也可定义关于 c 点的回差矩阵

$$D_y(s) \triangleq [I_m + H_{oy}(s)]$$

和 b 点的回差矩阵

$$D_u(s) \triangleq [I_l + H_{ou}(s)].$$

回差矩阵的维数分别与对应的回比矩阵的维数相同,可以证明

$$\det D_e(s) = \det D_u(s) = \det D_y(s).$$

对角优势矩阵(diagonal dominant matrix) 一类特殊的方阵. 复数域上的 $m\times m$ 矩阵 $A=[a_{ij}]$,如果满足

$$|a_{ii}| > \sum_{\substack{j=1\\(j\neq i)}}^{m} |a_{ij}| \triangleq d_i \quad (i=1,2,\cdots,m),$$

则称矩阵 A 为行对角优势矩阵. 第 i 行非对角元素的模之和 d_i 称为第 i 行的行估计或行半径. 若矩阵 A 满足:

$$|a_{ii}| > \sum_{\substack{j=1\\(j\neq i)}}^{m} |a_{ji}| \triangleq d'_i \quad (i=1,2,\cdots,m),$$

则称矩阵 A 为列对角优势矩阵. 第 i 列非对角元素的模之和 d'_i 称为第 i 列的列估计或列半径. 若矩阵 A 既为行对角优势矩阵,又为列对角优势矩阵,则称 A 是对角优势矩阵. 行(列)对角优势矩阵必为非奇异矩阵,且其逆矩阵的各对角元素的模在同列(行)中为最大. 如果控制系统的开环传递函数矩阵在虚轴和无穷大半径组成的包围整个右半复平面的闭合曲线上处处有行(列)对角优势矩阵,则闭环系统稳定性可以仅根据其对角元素进行判断.

盖尔斯哥利圆(Gershgorin circle) 一种矩阵特征值估计的图示方法. 用方阵的各行(列)的元素构造一系列圆,用于近似判定方阵的特征值在复平面上的位置. 具体构造方法如下:以 $m\times m$ 复数矩阵 $A=[a_{ij}]$ 的对角元素 $a_{ii}(i=1,2,\cdots,m)$ 在复平面上的点为圆心,以第 i 行(列)所有非对角元素的模之和为半径作圆,称为矩阵 A 的第 i 行(列)的行(列)盖尔斯哥利圆. 可以证明,一个复数矩阵的所有特征值都在该矩阵的诸盖尔斯哥利圆的并集内. 若矩阵的各元素是复数自变量的函数,则其诸盖尔斯哥利圆也随此自变量而变化,在复平面上描出一个带状区域,称为盖尔斯哥利带. 控制系统传递函数矩阵的盖尔斯哥利带常用于分析和设计多变量控制系统.

奥斯特洛夫斯基定理(Ostrovski theorem) 刻画对角优势矩阵性质的一个定理. 用以估计一个对角优势矩阵 $A=[a_{ij}]$ 的对角元素 a_{ii} 与它的逆矩阵 $A^{-1}=[\hat{a}_{ij}]$ 的相应对角元素的倒数 $1/\hat{a}_{ii}$ 之差的定理. 设 $A=[a_{ij}]$ 为行对角优势矩阵,定义

$$\theta_i \triangleq \sum_{\substack{j=1\\(j\neq i)}}^{m} \frac{|a_{ij}|}{|a_{ii}|} = \frac{d_i}{|a_{ii}|} < 1,$$

$$\varphi_i \triangleq \max_{\substack{k\\(k\neq i)}}\{\theta_k\},$$

式中 d_i 为矩阵 A 的第 i 行盖尔斯哥利圆半径(参见"盖尔斯哥利圆"),则有

$$\left| a_{ii} - \frac{1}{\hat{a}_{ii}} \right| \leqslant \varphi_i d_i < d_i.$$

令 $d_{oi}=\varphi_i d_i$,则称以 a_{ii} 为圆心,以 d_{oi} 为半径的圆为矩阵 A 的第 i 行奥斯特洛夫斯基圆. 对角优势矩阵的奥斯特洛夫斯基圆必位于相应的盖尔斯哥利圆的内部.

伪对角化(pseudo-diagonalization) 一种将矩阵化为对角优势阵的方法. 它是对给定的 $m\times m$ 复常数矩阵 $G=[g_{ij}]$,求一个实常数矩阵 $K=[k_{ij}]$,使之左乘(右乘)G 后所得的矩阵 $Q=KG(Q=GK)$ 具有尽可能强的行(列)对角优势的方法. 其基本思路如下:$Q=KG$ 的任一元素为

$$q_{rt} = \sum_{j=1}^{m} k_{rj} g_{jt} \quad (r,t = 1,2,\cdots,m).$$

为尽量加强第 r 行的对角优势,可以在 $|q_{rr}|=1$ 的约束条件下求

$$\sum_{j=1}^{m} |q_{rt}| = \min.$$

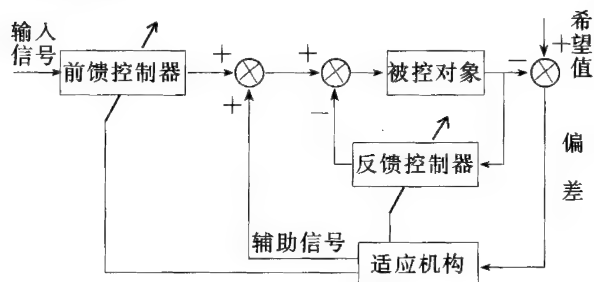
但这样做困难较大,故改为求

$$J_r = \sum_{\substack{j=1\\(t\neq r)}}^{m} |q_{rt}|^2 = \min \quad (r = 1,2,\cdots,m).$$

显然 J_r 只依赖于 K 的第 r 行而与 K 的其他行无关,因此这是一个 m 元函数的条件极值问题,可以用拉格朗日乘子法化为无条件极值问题求解. 依次令 $r=1,2,\cdots,m$,便可求得 K 的全部元. 汉夫金斯(Hawkins,D. J.)证明(1972),可以把这个问题归结为一个 $m\times m$ 矩阵的最大模特征值-特征向量问题求解. 后来约翰逊(Johnson,M. A.)又加以改进,归结为求解 2×2 矩阵的特征值-特征向量问题. 对于 $Q=GK$ 的情形,可用同样的方法进行伪对角化. 因为 $Q^T=K^TG^T$,故只需对 G 的转置 G^T 求 K 的转置 K^T,使 Q^T 有尽量强的行对角优势,则 Q 就有尽量强的列对角优势. 在多变量控制系统的设计中,常在某

一选定的频率 $s_1=j\omega_1$ 下，为被控制对象的传递函数矩阵 $G(s_1)$，用此法设计实常数预补偿器矩阵 K，使预补偿后的传递函数矩阵 $Q(s)$ 在 $s=s_1$ 点有尽可能强的对角优势，为应用逆奈奎斯特阵列设计方法创造条件.

自适应控制系统(adaptive control system) 一类具有一定适应能力的控制系统. 它是在一般控制系统的反馈回路中增加一个适应机构(控制器)的控制系统. 当被控对象的动态特性确切知道时，利用状态(或输出)与希望值之间的偏差形成的反馈控制器具有优良的控制精度和抗外干扰能力. 然而在实际系统中多数被控对象的动态特性是不大清楚的；或者虽然被控对象的动态特性清楚，但它在整个控制过程中变化较大. 在这些情况下，利用简单的反馈控制往往不能奏效. 为了获得较优良的系统性能品质和高精度，往往需要采取具有适应能力的自适应控制器. 所谓自适应控制器只是在一般控制系统的反馈回路中增加一个适应机构. 这个适应机构根据希望指标和实际指标的偏差随时修改控制器的参数或者产生一个辅助信号，使得闭环系统的性能指标尽可能接近所希望的性能指标. 自适应控制系统的一般结构见下图.



自适应控制系统结构图

模型参考自适应控制系统(model reference adaptive control system) 简称 MRACS. 自适应控制系统的一种形式. 它是设计适应机构使被控对象和已知参考模型的动态特性尽可能接近的一种自适应控制系统. 给定描述被控对象的希望动态特性的参考模型为

$$M:\ \begin{cases}\dot{x}_M = A_M x_M + B_M u,\\ y_M = C x_M.\end{cases}$$

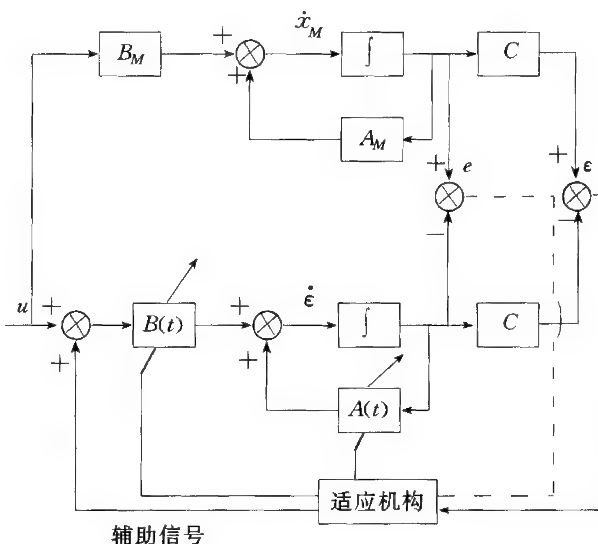
而被控对象(又称可调系统)的模型为

$$S:\ \begin{cases}\dot{x} = A(t)x + B(t)u,\\ y = Cx.\end{cases}$$

定义广义状态偏差 $e=x_M-x$ 和广义输出偏差 $\varepsilon=y_M-y$，以及由参考模型和被控对象的某些性能差构成的性能指标 J，通常记为

$$J = F(e,\varepsilon,A_M - A(t),B_M - B(t),t).$$

模型参考自适应控制器的作用机理是适应机构利用广义偏差而产生系统参数修正公式或产生辅助信号，使得性能指标函数 J 达到极小. 因此，性能指标最小化是设计适应机构的准则，而不要和参考模型所规定的系统希望动态特性相混淆. 模型参考自适应控制系统的一般结构见图.



模型参考自适应控制系统结构图

基于局部参数优化的自适应控制器(adaptive controller based on local parameter optimization) 一种自适应控制器. 它是在选定的指标下通过局部优化方法调整被控对象参数实现模型参考自适应控制的适应机构. 设参考模型为

$$\dot{x}_M = A_M x_M + B_M u,$$

而被控对象模型为

$$\dot{x} = A(e,t)x + B(e,t)u,$$

其中 $e=x_M-x$，$A(e,t)$，$B(e,t)$ 都是可调参数矩阵. 选择性能指标为

$$J = \frac{1}{2}e^{\mathrm{T}}e.$$

在假设可调参数的适应速度远远大于参数本身的变化速度的条件下，利用局部参数最优化思想，导出的可调参数矩阵的适应律为

$$\frac{\mathrm{d}A(e,t)}{\mathrm{d}t} = F_A e x^{\mathrm{T}},$$

$$\frac{\mathrm{d}B(e,t)}{\mathrm{d}t} = F_B e u^{\mathrm{T}},$$

其中 F_A 和 F_B 皆为 $n\times n$ 阶正定阵. 上式表明：基于局部参数优化的自适应控制器，用乘法器、加法器和积分器便可实现.

基于李亚普诺夫稳定性的自适应控制器(adaptive controller based on Liapunov stability) 一类由李亚普诺夫函数构造的自适应控制器. 它是通过李亚普诺夫稳定性方法使广义误差渐近趋于零而选定的模型参考自适应控制系统的适应机构. 设参考

模型为

$$\dot{x}_M = A_M x_M + B_M u,$$

而被控对象模型为

$$\dot{x} = A(e,t)x + B(e,t)u,$$

其中 $e=x_M-x$.

设 $\sigma(A_M)\subset \mathbb{C}^-$，其中 $\sigma(\cdot)$表示对应矩阵特征值集，$\mathbb{C}^-$系左半开复平面，则广义误差方程为

$$\dot{e}=A_M e+[A_M-A(e,t)]x+[B_M-B(e,t)]u.$$

所谓基于李亚普诺夫稳定性的自适应控制器，就是依据李亚普诺夫稳定性定理确定可调参数矩阵的适应律，使得

$$\lim_{t\to\infty} e(t) = 0.$$

由于 $\sigma(A_M)\subset \mathbb{C}^-$，据李亚普诺夫方程知道，对任意的 $Q>0$，必存在惟一 $P>0$，使得

$$PA_M + A_M^{\mathrm{T}}P = -Q.$$

用上式给定的 $P>0$ 构造如下的李亚普诺夫函数

$$V(e,t) = \frac{1}{2}e^{\mathrm{T}}Pe + \mathrm{tr}[A_M - A(e,t)]F_A^{-1}[A_M - A(e,t)] + \mathrm{tr}[B_M - B(e,t)]F_B^{-1}[B_M - B(e,t)],$$

其中 $F_A^{-1}>0$，$F_B^{-1}>0$，$\mathrm{tr}(\cdot)$表示矩阵的迹，即矩阵对角元素之和. 当取

$$\frac{\mathrm{d}A(e,t)}{\mathrm{d}t}=F_A Pex^{\mathrm{T}},\quad \frac{\mathrm{d}B(e,t)}{\mathrm{d}t}=F_B Peu^{\mathrm{T}},$$

便有

$$\frac{\mathrm{d}V(e,t)}{\mathrm{d}t} < 0,$$

因而有

$$\lim_{t\to\infty} e(t) = 0.$$

显然当 $P=I_n$ 时，该适应律就退化为基于局部参数优化的适应律.

基于超稳定性的自适应控制器(adaptive controller based on hyper stability)　根据超稳定性定理设计出的一种自适应控制器. 它是通过波波夫超稳定性定理和严格正实引理使广义误差渐近趋于零而选定的模型参考自适应控制系统的适应机构. 设参考模型为

$$\dot{x}_M=A_M x_M+B_M u_M,$$

而被控对象模型为

$$\dot{x} = Ax + Bu,$$

其中 A,B 在其标称值 A^0,B^0 处有摄动，即 A,B 是不确切知道的. 设 $\sigma(A_M)\subset \mathbb{C}^-$，即 A_M 是稳定阵，且 (A^0,B^0)和(A_M,B_M)都是能稳定的. 令 $e=x_M-x$，所谓设计自适应控制器，就是设计 u，不但使得当 $A=A^0,B=B^0$ 时，有

$$\lim_{t\to\infty} e(t)=0,$$

且当 A,B 在其标称值 A^0,B^0 处摄动时，仍有

$$\lim_{t\to\infty} e(t) = 0.$$

根据波波夫超稳定性定理和严格正实引理，只要取

$$u=K_M x_M+(K_u+\Delta K_u(e,t))u_M -(K+\Delta K(e,t))x, \tag{1}$$

其中 K_M 满足

$$\sigma(A_M - B^0K_M) \subset \mathbb{C}^-,$$
$$K = K_M + (B^0)^+ (A^0 - A_M),$$
$$K_u = (B^0)^+ B_M,$$

$(B^0)^+$是 B^0 的伪逆；而 $\Delta K(e,t)$，$\Delta K_u(e,t)$为

$$\begin{cases}\Delta K(e,t)=\int_0^t LDe(Hx)^{\mathrm{T}}\mathrm{d}\tau+\Delta K(0),\\ \Delta K_u(e,t)=\int_0^t MDe(Ru_M)^{\mathrm{T}}\mathrm{d}\tau+\Delta K_u(0),\end{cases} \tag{2}$$

或

$$\begin{cases}\Delta K(e,t) = \widetilde{L}De(\widetilde{H}x)^{\mathrm{T}},\\ \Delta K_u(e,t) = \widetilde{M}De(Ru_M)^{\mathrm{T}},\end{cases} \tag{3}$$

其中 $D=(B^0)^+P$，P 满足

$$P(A_M - B^0K_M) + (A_M - B^0K_M)^{\mathrm{T}}P = -Q,$$

且 $Q>0$，L,M,H,R 和 $\widetilde{L},\widetilde{M},\widetilde{H},\widetilde{R}$ 皆为正定阵. $\Delta K(0)$，$\Delta K_u(0)$为常阵，必有

$$\lim_{t\to\infty} e(t) = 0.$$

因此式(1)和(2)或(3)都是自适应控制器. 前者称为积分型，后者称为比例型.

切换系统(switching system)　一类具有模态切换的系统. 它在指定范围内按确定性或随机性规则改变系统动态模式的一类系统. 动态模式可表征为描述系统变量关系的一个微分方程. 切换系统属于特殊的一类时变系统，每个时间区间系统的状态方程和输出方程惟一，不同时间区间状态方程和输出方程为不同，切换时刻的状态满足连续性要求. 不同于一般的变结构控制，切换系统中切换模式序列和切换时刻序列都为可变因素，问题的复杂性要大得多. 目前，研究较多的限于线性切换系统，其动态模式可表示为

$$\begin{cases}\dot{x}(t)=A_{(m_i,t_i)}x(t)+B_{(m_i,t_i)}u(t),\\ y(t)=C_{(m_i,t_i)}x(t),\end{cases}$$

其中 x,y,u 为状态、输出和输入，系数矩阵$\{A_{(m_i,t_i)},B_{(m_i,t_i)},C_{(m_i,t_i)}\}$属于给定模式范围$\{A_k,B_k,C_k,k=1,2,\cdots,N\}$，$\{(m_i,t_i)\}_{i=0}^{\infty}$为切换序列，$m_i$ 属于正整数集合$\{1,2,\cdots,N\}$，t_i 为切换时刻. 主要研究问题涉及稳定性、能控性、镇定等，其他还有系统轨道的周期性、系统出现混沌的条件等.

切换系统的渐近稳定是指系统对所有可能的切换序列都使

$$\lim_{t\to\infty} \| x(t) \| = 0$$

成立. 主要研究手段是构造李亚普诺夫函数. 例如，

定常切换系统为稳定的一个充分条件是：对所有动态模式具有公共李亚普诺夫函数 $v(x)=x^{\mathrm{T}}Px$，其中正定对称矩阵 P 满足

$$A_i^{\mathrm{T}}P + PA_i < 0 \quad (i=1,2,\cdots,N).$$

线性切换系统的能控是指对任意非零初态 $x(t_0)=x_0$ 和任意目标状态 x_f，存在时刻 t_f 和切换序列，以及相应模式下的控制 $u(t)$，使 $x(t_f)=x_f$ 成立. 能控性的主要判据有格拉姆矩阵判据和秩判据. 格拉姆矩阵判据为：系统能控的充分必要条件是格拉姆矩阵

$$W(t_0,t_N) = \sum_{i=1}^{N}\int_{t_{i-1}}^{t_i}\Phi(t,t_0)B_iB_i^{\mathrm{T}}\Phi(t,t_0)^{\mathrm{T}}\mathrm{d}t$$

为非奇异，其中 $\Phi(t,t_0)=\Phi^{-1}(t_0,t)$ 即系统状态转移矩阵的逆. 秩判据为：系统能控的充分必要条件是对给定切换时刻 $0<t_1<\cdots<t_N$，判别矩阵

$$\tilde{C}=[C_1 \ \cdots \ \Phi(t_{i-1}t_0)C_i \ \cdots \ \Phi(t_{N-1},t_0)C_N]$$

为行满秩，其中

$$C_i=[B_i \ A_iB_i \ \cdots \ A_i^{n-1}B_i] \ (i=1,2,\cdots,N).$$

切换系统的镇定是指：通过选择模式、切换序列和相应模式的控制，使系统实现渐近稳定. 对能控的线性切换系统，必存在切换序列，以及状态反馈

$$u(t) = -K(t)x(t) + r(t)$$

和常数 $c_1>0$，使闭环状态转移矩阵为指数稳定：

$$\|\hat{\Phi}(t_0,t)\| \leqslant c_1\exp[-c_2(t-t_0)], \ t\geqslant t_0,$$

其中 $\|\cdot\|$ 为所示矩阵的谱范数，$c_2>0$ 为一常数，$K(t)$ 和 $r(t)$ 为反馈矩阵和参考输入.

奇异摄动控制系统

奇异摄动问题(singular perturbation problem) 一类摄动问题. 讨论带有小参数 $\varepsilon(0<\varepsilon\ll 1)$ 的微分方程的定解问题 P_ε. 这里小参数 ε 可以在方程中、初值或边值条件中、边界形状中出现. 如果该问题的解 $y_\varepsilon(x)$ 能表示成 ε 的幂级数：

$$y_\varepsilon(x) \simeq y_0(x) + \sum_{j=1}^{\infty}y_j(x)\varepsilon^j, \ x\in D, \ \varepsilon\to 0,$$

这里"$\simeq$"表示"渐近等于"，D 是问题 P_ε 的定义域，$y_0(x)$ 是问题 $P_\varepsilon|_{\varepsilon=0}$ 的解，且上述展开式对 $x\in\overline{D}$(D 的闭包)是一致收敛的，则称 P_ε 为正则摄动问题，否则，就称 P_ε 为奇异摄动问题. 这里，奇异摄动问题的范围很广，它包括问题 $P_0(\triangleq P_\varepsilon|_{\varepsilon=0})$ 无解；某一个解 $y_i(x)$ 不存在；上述渐近展开式不收敛；即使渐近展开式对每一点 $x\in D$ 均收敛，但在 $\overline{D}$ 内不一致收敛. 奇异摄动法在很多学科中有着重要的应用，如天文学、力学等. 在具有不同时间尺度的控制系统中，也有着其宽广的背景及重要的应用.

奇异摄动控制系统(singularly perturbed control system) 一类控制系统. 指奇异摄动的控制系统，其中小参数 ε 通常出现在该系统的高阶导数项的系数上，使整个系统呈现出具有高频、低频混合的特性. 常见的状态空间描述形式为

$$\begin{cases}\dot{x}=f(x,z,u,t,\varepsilon), \ x(t_0)=x_0,\\ \varepsilon\dot{z}=g(x,z,u,t,\varepsilon), \ z(t_0)=z_0,\end{cases}$$

这里 f 与 g 具有同阶的范数(或频谱). 对这一类系统，通常用边界层校正法求解，即把上述问题的解 (x,z) 分解成外解 (X,Z) 及边界层校正项 $(\overline{m},\overline{n})$，即

$$\begin{cases}x(t,\varepsilon) = X(t,\varepsilon) + \overline{m}(t,\varepsilon),\\ z(t,\varepsilon) = Z(t,\varepsilon) + \overline{n}(t,\varepsilon).\end{cases}$$

然后，将 $X,Z,\overline{m},\overline{n}$ 展成 ε 的幂级数：

$$X(t,\varepsilon) \simeq X_0(t) + \sum_{i=1}^{\infty}X_i(t)\varepsilon^i,$$

$$Z(t,\varepsilon) \simeq Z_0(t) + \sum_{i=1}^{\infty}Z_i(t)\varepsilon^i,$$

$$\overline{m}(t,\varepsilon) \simeq \overline{m}_0(t) + \sum_{i=1}^{\infty}\overline{m}_i(t)\varepsilon^i,$$

$$\overline{n}(t,\varepsilon) \simeq \overline{n}_0(t) + \sum_{i=1}^{\infty}\overline{n}_i(t)\varepsilon^i.$$

按边界层校正项 $\overline{m}_i$ 及 $\overline{n}_i$ 均仅在边界层内起作用的原理，可得到原问题具有各阶精度的近似解. 这里"$\simeq$"表示"渐近等于".

双时间尺度(double time scales) 奇异摄动系统的时标特性. 对于具有形如

$$\begin{cases}\dot{x} = f(x,z,t,\varepsilon), \ x(t_0) = x_0,\\ \varepsilon\dot{z} = g(x,z,t,\varepsilon), \ z(t_0) = z_0\end{cases}$$

的系统，由"奇异摄动控制系统"条目可知，它的解可表示成外解及边界层校正项之和

$$\begin{cases}x(t,\varepsilon) = X(t,\varepsilon) + \overline{m}(t,\varepsilon),\\ z(t,\varepsilon) = Z(t,\varepsilon) + \overline{n}(t,\varepsilon),\end{cases}$$

其中，边界层校正项是仅在很薄的一层边界层内起作用的，在边界层内，引入一种新的时间尺度 $\tau=(t-t_0)/\varepsilon$，记

$$m(\tau,\varepsilon) \triangleq \overline{m}(t_0+\varepsilon\tau,\varepsilon),$$
$$n(\tau,\varepsilon) = \overline{n}(t_0+\varepsilon\tau,\varepsilon),$$

则

$$\lim_{\tau\to\infty} m(\tau,\varepsilon) = 0, \ \lim_{\tau\to\infty} n(\tau,\varepsilon) = 0.$$

这样，对于原奇异摄动系统来说，该系统具有双时间尺度(t 与 τ)性质.

外解(outer solution) 组成奇异摄动问题解的一部分. 对奇异摄动系统

$$\begin{cases}\dot{x} = f(x,z,t,\varepsilon), \ x(t_0) = x_0,\\ \varepsilon\dot{z} = g(x,z,t,\varepsilon), \ z(t_0) = z_0\end{cases}$$

来说，其外解 $(X(t,\varepsilon),Z(t,\varepsilon))$ 满足方程

$$\begin{cases}\dot{X}(t,\varepsilon) = f(X,Z,t,\varepsilon),\\ \varepsilon\dot{Z}(t,\varepsilon) = g(X,Z,t,\varepsilon).\end{cases}$$

外解的适用区域在整个问题所讨论的时间区间内.而满足

$$\begin{cases}\dfrac{\mathrm{d}\overline{m}(\tau,\varepsilon)}{\mathrm{d}\tau}=\varepsilon[f(X(t,\varepsilon)+\overline{m}(\tau,\varepsilon),\\ \qquad Z(t,\varepsilon)+\overline{n}(\tau,\varepsilon),t,\varepsilon)-f(X(t,\varepsilon),\\ \qquad Z(t,\varepsilon),t,\varepsilon)]_{t=t_0+\varepsilon\tau},\\ \dfrac{\mathrm{d}\overline{n}(\tau,\varepsilon)}{\mathrm{d}\tau}=[g(X(t,\varepsilon)+\overline{m}(\tau,\varepsilon),\\ \qquad Z(t,\varepsilon)+\overline{n}(\tau,\varepsilon),t,\varepsilon)-g(X(t,\varepsilon),\\ \qquad Z(t,\varepsilon),t,\varepsilon)]_{t=t_0+\varepsilon\tau}\end{cases}$$

的$(\overline{m}(\tau,\varepsilon),\overline{n}(\tau,\varepsilon))$称为边界层校正项.它是组成奇异摄动问题解的另一部分.边界层校正项仅适用于边界层区域内.

在上面两组解的表达式中,按ε的幂级数展开可依次得到$\overline{m}_0(\tau)=0,X_0(t),Z_0(t),\overline{n}_0(\tau),\overline{m}_1(\tau),X_1(t),Z_1(t),\overline{n}_1(\tau),\overline{m}_2(\tau),\cdots$,对应的初始条件由相应的匹配性得到.

边界层校正项(boundary correcting term) 见"外解".

奇异摄动系统的分解(decomposition for singularly perturbed system) 将奇异摄动系统按高频、低频分解为子系统.对奇异摄动系统

$$\Sigma:\quad\begin{cases}\dot{x}=f(x,z,t,\varepsilon),\\ \varepsilon\dot{z}=g(x,z,t,\varepsilon),\end{cases}$$

如果方程$g(x,z,t,0)=0$对z有解,则称原系统在$\varepsilon=0$时的

$$\Sigma_s:\quad\begin{cases}\dot{x}_s=f(x_s,z_s,t,0),\\ g(x_s,z_s,t,0)=0\end{cases}$$

为该系统的降价系统,也称Σ_s为Σ的慢变子系统.引入$z_f\triangleq z-z_s$,可得关于$z_f(\tau)$所满足的方程

$$\Sigma_f:\quad\frac{\mathrm{d}z_f(\tau)}{\mathrm{d}\tau}=g(x_s,z_s+z_f,t_0,0).$$

称Σ_f为原系统的快变子系统.奇异摄动系统反映了双时间尺度的性质.一类时间尺度的特性由x_s及z_s反映出来,它表示系统的慢变的、低频的特征;另一类时间尺度的特性,则是表现快速的、高频的变化特征.又称Σ_s与Σ_f为原系统Σ的慢变及快变的分解.进一步分析可以看出,其慢变子系统的解(x_s,z_s)即是奇异摄动系统的外解(X,Z)关于ε幂级数展开式的零阶项(X_0,Z_0);而快变子系统的解$z_f(\tau)$是奇异摄动系统的边界层校正项$(\overline{m}(\tau),\overline{n}(\tau))$关于$\varepsilon$的幂级数展开式的零阶项$(\overline{m}_0,\overline{n}_0)$的一部分$\overline{n}_0(\tau)(\overline{m}_0(\tau)=0)$.因此,也称快变子系统为原系统的边界层系统.对于带有量测方程的系统

$$\begin{cases}\dot{x}=f(x,z,t,\varepsilon),\\ \varepsilon\dot{z}=g(x,z,t,\varepsilon),\\ y=h(x,z,t,\varepsilon),\end{cases}$$

称

$$\begin{cases}\dot{x}_s=f(x_s,z_s,t,0),\\ 0=g(x_s,z_s,t,0),\\ y_s=h(x_s,z_s,t,0)\end{cases}$$

及

$$\begin{cases}\dfrac{\mathrm{d}z_f(\tau)}{\mathrm{d}\tau}=g(x_s,z_s+z_f,t_0,0),\\ y_f=h(x_s,z_s+z_f,t_0,0)-h(x_s,z_s,t_0,0)\end{cases}$$

为其原系统的慢变子系统及快变子系统.

奇异摄动控制系统的组合控制(combined control for singularly perturbed control system) 一种针对奇异摄动控制系统的有效控制方案.对于奇异摄动控制系统

$$\begin{cases}\dot{x}=f(x,z,u,t,\varepsilon),\\ \varepsilon\dot{z}=g(x,z,u,t,\varepsilon),\end{cases}$$

它可以分解成相应的慢变与快变子系统

$$\begin{cases}\dot{x}_s=f(x_s,z_s,u_s,t,0),\\ 0=g(x_s,z_s,u_s,t,0)\end{cases}$$

及

$$\frac{\mathrm{d}z_f}{\mathrm{d}\tau}=g(x_s,z_s+z_f,u_s+u_f,t_0,0).$$

它们分别可设计出所需的控制律$u_s(x_s)$及$u_f(z_f)$.这里,由u_s及u_f,可叠加得组合控制u_c:

$$u_c(x,z)=u(x)+u_f(z-h(x,u(x))),$$

这里$z_s=h(x_s,u_s)$是方程$g(x_s,z_s,u_s,t,0)=0$的解.原系统在该组合控制作用下,可得到很好的控制效果.这样,就可把一个高阶的奇异摄动控制系统的设计问题,化为两个低阶的正常系统的设计问题来做.例如,对一个$(n+m)$阶的线性奇异摄动系统的二次性能指标的最优控制问题,化成两个阶数分别为n阶及m阶的正常系统的二次性能指标的最优调节问题后,由所得的两个最优反馈控制律运用上述方法得到的组合控制u_c与原问题的最优控制u^*仅相差小参数ε的一阶小量,即

$$u_c=u^*+O(\varepsilon).$$

而控制效果仅相差ε的一个二阶小量,即

$$J_c=J^*+O(\varepsilon^2),$$

这里J为二次性能指标的泛函值.

线性奇异摄动控制系统的强能控性(strong controllability of linear singularly perturbed control system) 线性奇异摄动控制系统特有的一类能控性.对线性定常奇异摄动控制系统

$$\Sigma:\quad\begin{cases}\dot{x}=A_{11}x+A_{12}z+B_1u,\\ \varepsilon\dot{z}=A_{21}x+A_{22}z+B_2u,\\ y=C_1x+C_2z,\end{cases}$$

如果A_{22}非异,则它可分解成相应的慢变子系统

$$\Sigma_s:\quad\begin{cases}\dot{x}_s=A_0x_s+B_0u_s,\\ y_s=C_0x+D_0u_s,\end{cases}$$

和快变子系统

$$\Sigma_f:\quad \begin{cases}\dfrac{dz_f}{d\tau}=A_{22}z_f+B_2u_f,\\ y_f=C_2z_f,\end{cases}$$

这里

$$A_0=A_{11}-A_{12}A_{22}^{-1}A_{21},$$
$$B_0=B_1-A_{12}A_{22}^{-1}B_2,$$
$$C_0=C_1-C_2A_{22}^{-1}A_{21},$$
$$D_0=-D_2A_{22}^{-1}B_2.$$

如果慢变子系统能控且快变子系统(即边界层系统)也能控,则称原系统是强能控的.类似地,如果 Σ_f 与 Σ_s 均能观,则称原系统是强能观的.对一个线性定常奇异摄动控制系统来说,该系统强能控(观)可推出该系统是能控(观)的,反之不然.因此,奇异摄动控制系统的强能控、强能观是比整个系统能控、能观更强的一类特性.

线性奇异摄动控制系统的强能观性(strong observability of linear singularly perturbed control system) 见"线性奇异摄动控制系统的能控性".

两个频率尺度的传递函数阵(transfer function matrix with two frequency scales) 简记为TFSTF阵.一类奇异摄动系统的频域特性.指满足下列四个条件的有理分式阵 $W(s,\varepsilon)$:

1. $W(s,\varepsilon)$关于 s 是正则的(指是 s 的真有理分式阵或严格真有理分式阵),且关于 s 的系数是 ε 的实函数,且在 $\varepsilon=0$ 处解析.

2. $W(s,0)$是确定的且关于 s 是正则的.

3. $W\left(\dfrac{p}{\varepsilon},\varepsilon\right)\Big|_{\varepsilon=0}$是确定的且关于 p 是正则的.

4. $W(s,\varepsilon)$的每个极点 $s(\varepsilon)$在 $\varepsilon=0$ 处的渐近展开式为下列两形式之一:

1) $s_i(\varepsilon)=\sum\limits_{k=0}^{\infty}b_k(\varepsilon^{\frac{1}{q}})^k.$

2) $s_j(\varepsilon)=\dfrac{1}{\varepsilon}\sum\limits_{k=0}^{\infty}b_k(\varepsilon^{\frac{1}{q}})^k\quad(b_0\neq0).$

这里,q 为一个非负数,$\varepsilon^{\frac{1}{q}}$满足 $\nu^q=\varepsilon$.

例 1:$W(s,\varepsilon)=\dfrac{1}{\varepsilon s^2-(1+2\varepsilon)s+1}\quad(0<\varepsilon\ll1)$是TFSTF 阵.

例 2:$W(s,\varepsilon)=\dfrac{(1+\varepsilon)s+2}{\varepsilon s^2-1}$不是 TFSTF 阵,它不符合条件 2 和条件 4.

例 3:

$$W(s,\varepsilon)=\begin{pmatrix}\dfrac{\varepsilon s-2}{(s-2)(\varepsilon s-1)} & \dfrac{1}{\varepsilon s-1}\\ \dfrac{\varepsilon s-2}{(s-1)(s-2)(\varepsilon s-1)} & \dfrac{s-2}{(s-1)(\varepsilon s-1)}\end{pmatrix}$$

是 TFSTF 阵.

TFSTF 阵的强既约实现(strong reduced realization for TFSTF matrix) 奇异摄动系统的一种特殊实现.一个 TFSTF 阵 $W(s,\varepsilon)$的强既约实现是指:一个带有小参数 ε 的控制系统 Σ_ε,它是既强能控、又强能观的,且它的传递函数阵$W_\varepsilon(s)$与 $W(s,\varepsilon)$仅相差一个 ε 的一阶小量,即

$$W_\varepsilon(s)=W(s,\varepsilon)+O(\varepsilon).$$

一个 TFSTF 阵 $W(s,\varepsilon)$的强既约实现是一个阶数不比 $W(s,\varepsilon)$的最小实现的阶数高的系统.由于 $W(s,\varepsilon)$的最小实现不一定是强既约的,因此,按此定义的强既约实现是一种具有更好性能的实现.特别地,一个 TFSTF 阵的强既约实现有可能不是一个奇异摄动控制系统.但人们也把这个强既约实现看成是一个"特殊"形式的奇异摄动控制系统.

奇异摄动系统的慢流形法(slow manifold method for singularly perturbed systems) 研究奇异摄动系统的一种有效方法.对于$(n+m)$ 阶奇异摄动系统

$$\Sigma_\varepsilon:\quad\begin{cases}\dot{x}=f(x,z,\varepsilon)\quad(x\in\mathsf{R}^n),\\ \varepsilon\dot{z}=g(x,z,\varepsilon)\quad(x\in\mathsf{R}^m,0<\varepsilon\ll1),\end{cases}$$

考察在 R^{n+m}空间中的一个低维流形 $M_\varepsilon:z=\varphi(x,\varepsilon)$.在该流形上,奇异摄动系统 Σ_ε 的精确解等价于满足条件

$$g(x,\varphi(x,\varepsilon),\varepsilon)=\varepsilon\cdot\frac{\partial\varphi}{\partial x}\cdot f(x,\varphi(x,\varepsilon),\varepsilon)\quad(1)$$

的一个 n 阶子系统 $\Sigma_x:\dot{x}=f(x,\varphi(x,\varepsilon),\varepsilon)$.这个流形 Σ_ε 是 n 维的,称 M_ε 为系统 Σ_ε 的慢流形.称条件(1)为相应于 Σ_ε 的慢流形法的匹配条件.Σ_ε 的慢流形法由下述过程完成:首先将慢流形 Σ_ε 表示成

$$M_\varepsilon:z=\varphi_0(x)+\varepsilon\varphi_1(x)+\varepsilon^2\varphi_2(x)+\cdots,$$

这样得到各阶近似流形

$$M_i:z=\varphi_0(x)+\varepsilon\varphi_1(x)+\cdots+\varepsilon^i\varphi_i(x)$$
$$(i=0,1,2,\cdots);$$

然后相应地求上述子系统 Σ_x 的各阶精度的近似解 $x_i(t)$;最后可得到 Σ_ε 在各阶近似流形 M_i 上各阶精度的近似解$(x_i(t),z_i(t))$.这里

$$\dot{x}_0=f(x_0,\varphi_0(x_0),0),$$

即是原系统 Σ_ε 的降阶子系统,其中 $\varphi_0(x_0)$是 $g(x_0,z_0,0)=0$ 的解 $z_0=\varphi_0(x)$.

奇异弧(singular arc) 奇异摄动系统的曲线特殊段.众所周知,线性系统

$$\dot{x}=Ax+Bu,\ x(0)=x_0$$

使二次性能指标

$$J=\frac{1}{2}\int_0^T[x^TQx+u^TRu]dt$$

取极小的最优控制问题,要求 $R>0$.当 R 仅是半正定或甚至于为零阵时,把问题转而求使

$$J_\varepsilon=\frac{1}{2}\int_0^T[x^TQx+\varepsilon^2u^TRu]dt\quad(0<\varepsilon\ll1)$$

取极小的最优控制问题，并进而讨论 $\varepsilon\to 0$ 的极限情形. 这时最优控制

$$u_\varepsilon^* = -\frac{1}{\varepsilon^2}R^{-1}B^{\mathrm{T}}p,$$

相当于对原系统一个高增益反馈，而 p,x 满足

$$\begin{cases}\varepsilon^2\dot{x}=\varepsilon^2Ax-BR^{-1}B^{\mathrm{T}}p, & x(0)=x_0,\\ \dot{p}=-Qx-A^{\mathrm{T}}p, & p(T)=0,\end{cases}$$

它是一个带有小参数 $\mu=\varepsilon^2$ 的奇异摄动系统，是一个非标准的奇异摄动系统. 其降阶系统

$$\begin{cases}0 = BR^{-1}B^{\mathrm{T}}p_s,\\ \dot{p}_s = -Qx_s - A^{\mathrm{T}}p_s, \quad p_s(T) = 0\end{cases}$$

的轨线就称为奇异弧，最优轨线的两部分由该奇异弧连结，进入及离开奇异弧，均需要脉冲控制的作用.

广义线性系统及其他

广义线性定常系统(singular linear time-invariant system)　状态变量导数不能由显式解出的线性定常控制系统. 亦称描述变量系统，或广义状态空间系统、半状态系统、微分代数系统、奇异的奇异摄动系统、退化系统、受限系统等，它的一般状态空间形式为

$$\Sigma:\quad \begin{cases}E\dot{x} = Ax + Bu,\\ y = Cx.\end{cases}$$

简记为 $\Sigma(E,A,B,C)$. 其中 $x\in\mathrm{R}^n$ 是状态(亦有另外一种看法把 Ex 作为状态)，$u\in\mathrm{R}^r$ 是控制，$y\in\mathrm{R}^m$ 是输出. E,A,B,C 是适当维常数阵，且 $\mathrm{rank}\,E=q\leqslant n$，$q$ 称为广义阶；$\mathrm{rank}\,E=n$ 时，广义线性定常系统成为正常线性定常系统. 记 E 的幂零指数为 $\mathrm{Ind}(E)=h$，C_p^{h-1} 为具有 $h-1$ 阶分段连续可微函数集合. 对任一 $u\in C_p^{h-1}$，$\Sigma(E,A,B,C)$ 有惟一解的充分必要条件是 $\det(sE-A)\neq 0$(即 (E,A) 为正则束，简称 $\Sigma(E,A,B,C)$ 为正则的). 显然，被讨论的广义线性定常系统都应是正则的. 关于正则广义线性定常系统的求解有两种方法：一种是利用矩阵束理论，另一种是利用矩阵的 Drazin 逆.

为了书写方便，今后将广义线性定常系统简称广义系统.

Drazin 逆(Drazin inverse)　方阵逆概念对长阵的一种推广. 给定 $A\in\mathrm{C}^{n\times n}$，称使

$$\mathrm{Ker}(A^\nu) = \mathrm{Ker}(A^{\nu+1})$$

成立的最小非负整数 ν 为 A 的幂零指数，记为 $\mathrm{Ind}(A)=\nu$. 对任给的 $A\in\mathrm{C}^{n\times n}$ 总存在满秩方阵 T 使

$$A = T\begin{pmatrix}C & 0\\ 0 & N\end{pmatrix}T^{-1}, \tag{1}$$

其中 $C\in\mathrm{C}^{n_1\times n_1}$，$N\in\mathrm{C}^{n_2\times n_2}$，$C$ 为满秩方阵，N 为幂零阵，且 $n_1+n_2=n$. 若 $A\in\mathrm{C}^{n\times n}$ 且其表达式如(1)，记 A 的 Drazin 逆为 A^{D}，则

$$A^{\mathrm{D}}=T\begin{pmatrix}C^{-1} & 0\\ 0 & 0\end{pmatrix}T^{-1}.$$

显然 A^{D} 是惟一的，且有性质

$$AA^{\mathrm{D}}=A^{\mathrm{D}}A,$$
$$A^{\mathrm{D}}AA^{\mathrm{D}}=A^{\mathrm{D}},$$
$$A^{k+1}A^{\mathrm{D}}=A^k\quad (k\geqslant \mathrm{Ind}(A)).$$

广义系统的受限等价(restricted equivalence for singular system)　两个广义线性定常系统在相抵变换下的关系. 给定连续两个广义系统 $\Sigma(E,A,B,C)$ 和 $\bar{\Sigma}(\bar{E},\bar{A},\bar{B},\bar{C})$，如果存在两个满秩方阵 $P,Q\in\mathrm{R}^{n\times n}$，使

$$\bar{E} = QEP,\ \bar{A} = QAP,\ \bar{B} = QB,\ \bar{C} = CP,$$

则称 $\Sigma(E,A,B,C)$ 和 $\bar{\Sigma}(\bar{E},\bar{A},\bar{B},\bar{C})$ 是受限等价的.

给定多项式矩阵 $W(s)\in\mathrm{R}^{m\times r}(s)$ 和 $\bar{W}(s)\in\mathrm{R}^{m\times r}(s)$，如果存在单位模阵 $U(s)\in\mathrm{R}^{m\times m}(s)$，$V(s)\in\mathrm{R}^{r\times r}(s)$，使 $\bar{W}(s)=U(s)W(s)V(s)$，则称 $W(s)$ 和 $\bar{W}(s)$ 是等价的，将 $\bar{E}=QEP$ 和 $\bar{A}=QAP$ 写成多项式矩阵形式为

$$(s\bar{E} - \bar{A}) = Q(sE - A)P,$$

上式作为两个多项式矩阵 $(s\bar{E}-\bar{A})$ 和 $(sE-A)$ 的等价变换，相当于取 $U(s)=Q$，$V(s)=P$，即将两个单位模阵限制取为满秩方阵. 受限等价变换中的“受限”源于此.

广义系统的标准快分解(standard fast decomposition for singular system)　广义线性定常系统在相抵变换下的一种标准形式. 给定正则广义系统 $\Sigma(E,A,B,C)$，据正则矩阵束理论必存在满秩方阵 $Q\in\mathrm{R}^{n\times n}$，$P\in\mathrm{R}^{n\times n}$，使得

$$QEP=\begin{bmatrix}I_{n_1} & 0\\ 0 & N\end{bmatrix},\ QAP=\begin{bmatrix}A_1 & 0\\ 0 & I_{n_2}\end{bmatrix},$$

其中 $n_1+n_2=n$，N 是幂零阵.

只要取 $P^{-1}x=[x_1^{\mathrm{T}},x_2^{\mathrm{T}}]^{\mathrm{T}}$，$x_1\in\mathrm{R}^{n_1}$，$x_2\in\mathrm{R}^{n_2}$，$QB=[B_1^{\mathrm{T}},B_2^{\mathrm{T}}]^{\mathrm{T}}$，$CP=[C_1,C_2]$，则 $\Sigma(E,A,B,C)$ 受限等价于

$$\Sigma^*:\quad \begin{cases}\dot{x}_1 = A_1x_1 + B_1u,\\ N\dot{x}_2 = x_2 + B_2u.\end{cases}$$

即将广义系统分解为两个状态 x_1,x_2 解耦的子系统，它们是通过控制 u 耦合在一起的. 第一个子系统(对应于状态 x_1)相当于正常系统，它对应的系统极点为有限的，称为慢变子系统；第二个子系统(对应于状态 x_2)对应的极点是无穷的，称为快变子系统. Σ^* 称为广义系统 $\Sigma(E,A,B,C)$ 的标准快、慢分解.

广义系统的标准慢分解(standard slow decomposition for singular system)　见“广义系统的标准快分解”.

相容初始条件(consistent initial condition)

广义系统满足一定约束的初始条件. 它满足状态和控制的诸阶导数在"0_+"处的约束关系. 关于广义系统的解有两种看法:一种是容许初值及其对应的解,另一种是任意初值 $\bar{x}_0 \in \mathbf{R}^n$ 及其对应的解. 记 $P^{-1}x_0 = [x_{10}^{\mathrm{T}}, x_{20}^{\mathrm{T}}]^{\mathrm{T}}$. 对任一控制 $u(t) \in C_p^{h-1}$,通过标准快变、慢变子系统,可得其解为

$$x(t)=P\begin{bmatrix}I_{n_1}\\0\end{bmatrix}\{\mathrm{e}^{A_1t}[I_{n_1},0]P^{-1}x_0 + \int_0^t \mathrm{e}^{A_1(t-\tau)}B_1u(\tau)\mathrm{d}\tau\} - P\begin{bmatrix}0\\I_{n_2}\end{bmatrix}\sum_{i=0}^{h-1}N^iB_2u^{(i)}(t), \quad t>0,$$

其中 $u^{(i)}$ 是 u 的 i 阶导数. 当 $t \to 0_+$ 时,由上式得

$$x(0_+) = P\begin{bmatrix}I_{n_1}\\0\end{bmatrix}[I_{n_1},0]P^{-1}x_0 - P\begin{bmatrix}0\\I_{n_2}\end{bmatrix}\sum_{i=0}^{h-1}N^iB_2u^{(i)}(0_+).$$

满足上述条件的初值,称为容许初值. 它是加在初值上的一种限制,而这种限制与控制 $u^{(i)}(0_+)$,$i=0,1,2,\cdots,h-1$ 的选择有关.

广义系统的分布解(distribution solution of singular system) 广义系统的一种解. 将状态和控制都视为广义函数时广义线性定常系统的解. 前面介绍了容许初值及其解,且容许初值是依赖于控制及其诸阶导数在初始时刻的值. 但由于在控制系统中,控制总是根据对其系统性能品质的要求,从容许控制集合中挑选的,因此容许初值对某些控制问题的讨论是不方便的. 为了讨论广义系统解的性质,人们希望广义系统的初值能像正常系统那样能在 $\mathbf{R}^n$ 中任取. 将状态和控制都视为广义函数,在分布意义下获得了对任意初始条件 x_0 的解

$$x(t)=P\begin{bmatrix}I_{n_1}\\0\end{bmatrix}\{\mathrm{e}^{A_1t}[I_{n_1},0]P^{-1}x_0 + \int_0^t \mathrm{e}^{A_1(t-\tau)}B_1u(\tau)\mathrm{d}\tau\} - P\begin{bmatrix}0\\I_{n_2}\end{bmatrix}\{\sum_{i=1}^{h}\delta^{(i-1)}(t)N^i[0,I_{n_2}]P^{-1}x_0 + \sum_{i=0}^{h-1}N^iB_2u^{(i)}(t)\},$$

其中 $\delta^{(j)}(t)$ 是 δ 函数的第 j 阶导数. 比较上式和容许初值的解知,其差别在于后者在初始时刻($t_0=0$)处多了一些脉冲项. 正是由于这些脉冲项的存在,才能把任意初始条件瞬时驱动成 0_+ 时刻的容许初值,当 $t>0$ 时两个解完全相同.

广义系统的极点(pole of singular system) 与广义系统性质相关的一个重要概念. 广义系统所对应的广义特征方程的根. 给定广义系统 $\Sigma(E,A,B,C)$,记 $\sigma(E,A)=\{s \mid \det(sE-A)=0\}$,则称 $\sigma(E,A)$ 为广义系统的有限极点集合. 其个数为

$$\deg\det(sE-A) \leqslant \operatorname{rank}E < n.$$

如果将 x 视为状态,$\Sigma(E,A,B,C)$ 为 n 阶系统,它应有 n 个极点,除 $\deg\det(sE-A)$ 个有限极点外,尚应有 $n-\deg\det(sE-A)$ 个无穷极点;如果把 Ex 视为状态,则 $\Sigma(E,A,B,C)$ 为 $\operatorname{rank}E$ 阶系统,除 $\deg\det(sE-A)$ 个有限极点外,尚应有 $\operatorname{rank}E-\deg\det(sE-A)$ 个无穷极点. 因此无论如何看待广义系统的状态,一般说,它都存在着无穷极点. 从广义系统的标准分解易知:有限极点对应于慢变子系统,而无穷极点则和快变子系统对应. 不失一般性,可设 N 具有若当块型. 如果把 Ex 视为状态,无穷远极点个数恰为 N 中非零元素的个数,而零若当块不对应着无穷极点. 如果把 x 视为状态,不管 N 的若当块是否为零皆对应于无穷极点,且其个数为 n_2.

广义系统的传递函数阵(transfer function matrix of singular system) 广义线性定常系统的一种频域描述. 给定广义系统 $\Sigma(E,A,B,C)$,则称

$$G(s) = C(sE-A)^{-1}B$$

为其传递函数阵. 由于受限等价不改变传递函数阵,因此,从 Σ 的受限等价系统直接得

$$\begin{aligned}G(s) &= C(sE-A)^{-1}B \\ &= C_1(sI_{n_1}-A_1)^{-1}B_1 + C_2(sN-I)^{-1}B_2 \\ &= C_1(sI_{n_1}-A_1)^{-1}B_1 - \sum_{i=0}^{h-1}s^iC_2N^iB_2.\end{aligned}$$

显然 $G(s)$ 是一非真有理分式阵. 从经典控制理论知:非真有理分式阵,反映了系统的非因果性,即系统任一瞬时的性质,除依赖于过去的控制信息还依赖于未来的控制信息. 它和广义系统的状态 $x(t)$ 除依赖于 $u(t)$ 外,还依赖 $u^{(i)}(t)(i=1,2,\cdots,h-1)$ 是相一致的. 从而知广义系统可以反映系统的非因果性质.

广义系统的可达集(reachable set of singular system) 描述广义线性定常系统控制对状态作用能力的基本概念. 设广义系统具有标准分解

$$\dot{x}_1 = A_1x_1 + B_1u_1,$$
$$N\dot{x}_2 = x_2 + B_2u_1,$$

其中 $x_1 \in \mathbf{R}^{n_1}$,$x_2 \in \mathbf{R}^{n_2}$,$\operatorname{Ind}(N)=h$.

任给 $w \in \mathbf{R}^n$,如果存在初值 $x_1(0)$,允许控制 $u \in C_p^{h-1}$ 和 $t_1>0$,使

$$x(t_1) = \begin{pmatrix}x_1(t_1)\\x_2(t_1)\end{pmatrix} = w,$$

则 w 称为可达的. 记 $R(x_1(0))$ 为从初值 $x_1(0)$ 出发的可达状态集,则 $R(x_1(0))=\{w \in \mathbf{R}^n \mid 存在\ u \in C_p^{h-1}\ 和\ t_1>0, 使\ x(t_1)=w\}$. 记 R 为广义系统的可

达集，则

$$R = \bigcup_{x_1(0)\in \mathbf{R}^{n_1}} R(x_1(0)).$$

直接计算得知

$$R = R_s \oplus \operatorname{Im}(N, NB_2, \cdots, N^{h-1}B_2),$$

这里 R_s 为慢变子系统的可达集.

广义系统R能控(R-controllability of singular system) 刻画广义线性定常系统结构性质的基本概念. 给定广义系统 $E\dot{x}=Ax+Bu$，如前记

$$[x_1^{\mathrm{T}}, x_2^{\mathrm{T}}]^{\mathrm{T}} = P^{-1}x,\ x_1 \in \mathbf{R}^{n_1},\ x_2 \in \mathbf{R}^{n_2},$$

如果对任给的 $t_1>0$ 和

$$[I_{n_1}, 0]P^{-1}x_0 = x_{10} \in \mathbf{R}^{n_1},\ w \in \mathbf{R}^{n},$$

总存在着 $u\in C_p^{h-1}$，使得在这个控制作用下有 $x(t_1)=w$，则称广义系统是 R 能控的. 它和广义系统在其可达集中能控是等价的，它实质上刻画了慢变子系统的能控性. 如果对任给的 $t_1>0$ 及 $x(0)\in \mathbf{R}^n$ 和 $w\in \mathbf{R}^n$，总存在容许控制 $u(t)\in C_p^{h-1}$，使得在这个控制作用下有 $x(t_1)=w$，则称广义系统是能控的.

广义系统能控(controllability of singular system) 见"广义系统 R 能控".

广义系统的脉冲能控(impulse controllability of singular system) 刻画广义线性定常系统结构性质的一个基本概念. 广义系统的 R 能控和能控概念都没有涉及广义系统的脉冲动态行为. 但从广义系统分布意义下的解知道，广义系统的状态响应中还包含脉冲部分. 这些脉冲可能由初始条件产生，亦可能由控制 $u^{(i)}(t)\ (i=0,1,2,\cdots,h-1)$ 的跳跃行为产生. 记 $x_\tau(t)$ 表示状态 $x(t)$ 在 τ 时刻的脉冲部分，由其分布解知

$$x_\tau(t) = -P\begin{pmatrix}0\\ I_{n_2}\end{pmatrix}\Big\{\sum_{i=1}^{h}\delta^{(i-1)}(t)N^i[0, I_{n_2}]P^{-1}x_0 + \sum_{i=1}^{h-1}N^iB_2\Delta_\tau u^{(i)}(t)\Big\},$$

其中 $\Delta_\tau u^{(i)}(t)$ 是 $u^{(i)}(t)$ 在 τ 时刻的脉冲. 从上式易知，任意 $w\in\mathbf{R}^n$ 作为 τ 时刻初值在 τ 时刻产生的脉冲部分为

$$I_\tau(w,t) = -P\begin{pmatrix}0\\ I_{n_2}\end{pmatrix}\Big\{\sum_{i=1}^{h}\delta^{(i-1)}(t-\tau) \cdot N^i[0, I_{n_2}]P^{-1}w\Big\}.$$

如果对任意的 $x_0\in\mathbf{R}^n, w\in\mathbf{R}^n, \tau\in\mathbf{R}^n$，总存在容许控制 $u(t)\in C_p^{h-1}$，使得 $x_\tau(t)=I_\tau(w,t)$，则称广义系统为脉冲能控的. 该定义类似于正常动态行为的能控定义. 它刻画了容许控制产生脉冲的能力.

广义系统的强能控(strong controllability of singular system) 描述广义线性定常系统结构性质的基本概念之一. 给定广义系统

$$\Sigma: E\dot{x} = Ax + Bu,$$

如果广义系统 Σ 既是 R 能控的，又是脉冲能控的，则称广义系统 Σ 是强能控的. 广义系统的强能控概念实际上是把 Ex 作为状态(而不是把 x 作为状态)的普通意义下的能控性.

广义系统的能稳(stabilizability of singular system) 广义线性定常系统的两种结构性质. 给定广义系统 $\Sigma(E,A,B,C)$，如果其标准分解中慢变子系统是能稳的，即慢变子系统中的不能控的有限极点是稳定的，则称广义系统 $\Sigma(E,A,B,C)$ 是能稳的；如果广义系统 $\Sigma(E,A,B,C)$ 既是能稳的，又是脉冲能控的，则称广义系统 $\Sigma(E,A,B,C)$ 是强能稳的.

注：广义系统 $\Sigma(E,A,B,C)$ 的稳定是指李亚普诺夫意义下的渐近稳定.

广义系统的强能稳(strong stabilizability of singular system) 见"广义系统的能稳".

广义系统能控性判据(controllability criterion for singular system) 判别广义线性定常系统各种能控性的条件和方法. 给定 n 阶广义系统

$$\Sigma:\quad E\dot{x} = Ax + Bu.$$

1. 广义系统 Σ 能控的充分必要条件为

   $\operatorname{rank}(E,B)=n$,

   $\operatorname{rank}[sE-A, B]=n,\ \forall s\in \mathbf{C}^+$ 有限.

2. 广义系统 Σ R 能控的充分必要条件为

   $\operatorname{rank}[sE-A, B]=n,\ \forall s\in \mathbf{C}^+$ 有限.

3. 广义系统 Σ 脉冲能控的充分必要条件为

$$\operatorname{rank}\begin{pmatrix}E & 0 & 0\\ A & E & B\end{pmatrix} = n + \operatorname{rank} E.$$

4. 广义系统 Σ 强能控的充分必要条件为

   $\operatorname{rank}[sE-A, B]=n,\ \forall s\in \mathbf{C}^+$ 有限,

$$\operatorname{rank}\begin{pmatrix}E & 0 & 0\\ A & E & B\end{pmatrix} = n.$$

注：$\mathbf{C}^+$ 有限表示不含无穷远点的右半闭复平面.

广义系统的能观性(observability of singular system) 描述广义线性定常系统观测能力的重要概念. 给定广义系统

$$\Sigma:\quad \begin{cases}E\dot{x} = Ax + Bu,\\ y = Cx,\end{cases}$$

则称

$$\Sigma^{\mathrm{T}}:\quad \begin{cases}E^{\mathrm{T}}\dot{\xi} = A^{\mathrm{T}}\xi + C^{\mathrm{T}}v,\\ \eta = B^{\mathrm{T}}\xi\end{cases}$$

为 $\Sigma(E,A,B,C)$ 的对偶系统(显然 Σ 和 Σ^{T} 彼此为对偶系统).

1. 如果 $\Sigma^{\mathrm{T}}(E^{\mathrm{T}}, A^{\mathrm{T}}, C^{\mathrm{T}}, B^{\mathrm{T}})$ 是能控的，则称 $\Sigma(E,A,B,C)$ 为能观的.

2. 如果 $\Sigma^{\mathrm{T}}(E^{\mathrm{T}}, A^{\mathrm{T}}, C^{\mathrm{T}}, B^{\mathrm{T}})$ 是 R 能控的，则称广义系统 $\Sigma(E,A,B,C)$ 是 R 能观的.

3. 如果 $\Sigma^{\mathrm{T}}(E^{\mathrm{T}}, A^{\mathrm{T}}, C^{\mathrm{T}}, B^{\mathrm{T}})$ 是脉冲能控的，则称

$\Sigma(E,A,B,C)$是脉冲能观的.

4. 如果$\Sigma^{\mathrm{T}}(E^{\mathrm{T}},A^{\mathrm{T}},C^{\mathrm{T}},B^{\mathrm{T}})$是强能控的，则称$\Sigma(E,A,B,C)$为强能观的.

广义系统的能检测(detectability of singular system) 广义线性定常系统的一种结构性质. 给定广义系统$\Sigma(E,A,B,C)$，

1. 如果对偶系统$\Sigma^{\mathrm{T}}(E^{\mathrm{T}},A^{\mathrm{T}},C^{\mathrm{T}},B^{\mathrm{T}})$是能稳的，则称广义系统$\Sigma(E,A,B,C)$是能检测的.

2. 如果对偶系统$\Sigma^{\mathrm{T}}(E^{\mathrm{T}},A^{\mathrm{T}},C^{\mathrm{T}},B^{\mathrm{T}})$是强能稳的，则称广义系统$\Sigma(E,A,B,C)$是强能检测的.

广义系统的强能检测(strong detectability of singular system) 见“广义系统的能检测”.

广义系统的标准结构(canonical structure of singular system) 广义线性定常系统在相抵变换下的一种标准形式. 给定广义系统

$$\Sigma:\quad \begin{cases} E\dot{x} = Ax + Bu, \\ y = Cx, \end{cases}$$

则存在满秩阵$Q\in\mathbb{R}^{n\times n}$，$P\in\mathbb{R}^{n\times n}$，使得$\Sigma(E,A,B,C)$受限等价于

$$\widetilde{\Sigma}:\quad \begin{cases} \widetilde{E}\dot{\xi}=\widetilde{A}\xi+\widetilde{B}v, \\ \eta=\widetilde{C}\xi, \end{cases}$$

其中

$$\widetilde{E}=\begin{pmatrix} \widetilde{E}_{11} & 0 & \widetilde{E}_{13} & 0 \\ \widetilde{E}_{21} & \widetilde{E}_{22} & \widetilde{E}_{23} & \widetilde{E}_{24} \\ 0 & 0 & \widetilde{E}_{33} & 0 \\ 0 & 0 & \widetilde{E}_{43} & \widetilde{E}_{44} \end{pmatrix},$$

$$\widetilde{A}=\begin{pmatrix} \widetilde{A}_{11} & 0 & \widetilde{A}_{13} & 0 \\ \widetilde{A}_{21} & \widetilde{A}_{22} & \widetilde{A}_{23} & \widetilde{A}_{24} \\ 0 & 0 & \widetilde{A}_{33} & 0 \\ 0 & 0 & \widetilde{A}_{43} & \widetilde{A}_{44} \end{pmatrix},$$

$$\widetilde{B}=[\widetilde{B}_{11}^{\mathrm{T}},\widetilde{B}_{12}^{\mathrm{T}},0,0]^{\mathrm{T}},$$

$$\widetilde{C}=[\widetilde{C}_{11},0,\widetilde{C}_{13},0].$$

且有

1. 子系统$\Sigma_1(\widetilde{E}_{11},\widetilde{A}_{11},\widetilde{B}_{11},\widetilde{C}_{11})$既能控又能观测.

2. 子系统

$$\Sigma_2\left(\begin{pmatrix} \widetilde{E}_{11} & 0 \\ \widetilde{E}_{21} & \widetilde{E}_{22} \end{pmatrix},\begin{pmatrix} \widetilde{A}_{11} & 0 \\ \widetilde{A}_{21} & \widetilde{A}_{22} \end{pmatrix},\begin{pmatrix} \widetilde{B}_{11} \\ \widetilde{B}_{12} \end{pmatrix},[\widetilde{C}_{11},0]\right)$$

是能控但不能观测的.

3. 子系统

$$\Sigma_3\left(\begin{pmatrix} \widetilde{E}_{11} & \widetilde{E}_{13} \\ 0 & \widetilde{E}_{33} \end{pmatrix},\begin{pmatrix} \widetilde{A}_{11} & \widetilde{A}_{13} \\ 0 & \widetilde{A}_{33} \end{pmatrix},\begin{pmatrix} \widetilde{B}_{1} \\ 0 \end{pmatrix},[\widetilde{C}_{11},\widetilde{C}_{13}]\right)$$

是能观测但不能控的.

4. 子系统$\Sigma_4(\widetilde{E}_{44},\widetilde{A}_{44},0,0)$是既不能控又不能观测的.

$\widetilde{\Sigma}(\widetilde{E},\widetilde{A},\widetilde{B},\widetilde{C})$称为广义系统$\Sigma$的标准结构.

广义系统的能观测性判据(observability criterion for singular system) 广义线性定常系统各种能观测性的判别方法. 据广义系统能观测性定义和能控性判别条件直接得：

1. 广义系统$\Sigma(E,A,B,C)$能观测的充分必要条件为$\mathrm{rank}[E^{\mathrm{T}},C^{\mathrm{T}}]^{\mathrm{T}}=n$；$\mathrm{rank}[sE^{\mathrm{T}}-A^{\mathrm{T}},C^{\mathrm{T}}]^{\mathrm{T}}=n$，$\forall s\in\mathbb{C}^+$有限.

2. 广义系统$\Sigma(E,A,B,C)R$能观测的充分必要条件为$\mathrm{rank}[sE^{\mathrm{T}}-A^{\mathrm{T}},C^{\mathrm{T}}]^{\mathrm{T}}=n$，$\forall s\in\mathbb{C}^+$有限.

3. 广义系统$\Sigma(E,A,B,C)$脉冲能观的充分必要条件为

$$\mathrm{rank}\begin{pmatrix} E^{\mathrm{T}} & 0 & 0 \\ A^{\mathrm{T}} & E^{\mathrm{T}} & C^{\mathrm{T}} \end{pmatrix}=n+\mathrm{rank}\ E.$$

4. 广义系统$\Sigma(E,A,B,C)$强能观的充分必要条件为

$$\mathrm{rank}[sE^{\mathrm{T}}-A^{\mathrm{T}},C^{\mathrm{T}}]=n,\ \forall s\in\mathbb{C}^+\text{有限；}$$

$$\mathrm{rank}\begin{pmatrix} E^{\mathrm{T}} & 0 & 0 \\ A^{\mathrm{T}} & E^{\mathrm{T}} & C^{\mathrm{T}} \end{pmatrix}=n+\mathrm{rank}\,E.$$

广义系统的最小实现(minimal realization of singular system) 广义系统从频域描述到时域描述的一种转换. 它是由广义线性定常系统的传递函数描述对应阶数最低的状态空间描述. 对给定的非真有理分式阵$G(s)\in\mathbb{R}^{m\times r}(s)$，如果存在广义系统$\Sigma(E,A,B,C)$，使得

$$G(s) = C(sE - A)^{-1}B,$$

其中$E\in\mathbb{R}^{n\times n}$，则称$\Sigma(E,A,B,C)$是$G(s)$的一个$n$阶广义状态空间实现，简称实现. 在$G(s)$的所有实现中阶数最低者称为$G(s)$的最小实现. 从广义系统的标准结构知道，$\Sigma(E,A,B,C)$为$G(s)$的最小实现的充分必要条件是$\Sigma(E,A,B,C)$既能控又能观测，且$G(s)$的两个最小实现必是受限等价的. 这就说明，在受限等价意义下，$G(s)$的最小实现是惟一的. 任给一非真有理分式阵$G(s)$，总存在如下分解

$$G(s) = G_{\mathrm{sp}}(s) + M(s),$$

其中$G_{\mathrm{sp}}(s)$是严格真有理分式阵，$M(s)$是多项式阵. 记$\Sigma_1(I_{n_1},A_1,B_1,C_1)$是严格真有理分式阵$G_{\mathrm{sp}}(s)$的最小实现(它可由线性系统理论得到). 剩下的问题是需要求得$M(s)$的最小实现. 记

$$W(\lambda) = -\frac{1}{\lambda}M\left(\frac{1}{\lambda}\right),$$

则$W(\lambda)$是λ的严格真有理分式阵. 如果$\Sigma_2(J,I_{n_2},B_2,C_2)$是$W(\lambda)$的最小实现，则$\Sigma(I_{n_2},J,B_2,C_2)$是$M(s)$的最小实现. 直接计算知

$$\Sigma\left(\begin{pmatrix} I_{n_1} & 0 \\ 0 & J \end{pmatrix},\begin{pmatrix} A_1 & 0 \\ 0 & I_{n_2} \end{pmatrix},\begin{pmatrix} B_1 \\ B_2 \end{pmatrix},[C_1,C_2]\right)$$

是非真有理分式阵$G(s)$的最小实现.

纯状态反馈(pure state feedback) 广义系统的一种反馈形式. 它是仅含有系统状态变量的反馈.

给定广义系统

$$\Sigma:\quad \begin{cases} E\dot{x} = Ax + Bu, \\ y = Cx. \end{cases}$$

形如 $u=Kx+v$(其中 $K\in\mathbb{R}^{r\times n}$)的反馈称为纯状态反馈，简称状态反馈. 在纯状态反馈下闭环系统为

$$\Sigma_C^{(1)}:\quad \begin{cases} E\dot{x} = (A+BK)x + Bv, \\ y = Cx \end{cases}$$

闭环系统 $\Sigma_C(E,A+BK,B,C)$ 具有如下性质：

1. 对任意的反馈增益阵 $K\in\mathbb{R}^{r\times n}$，闭环系统至多有 $\operatorname{rank}E$ 个有限极点.

2. 存在一个 $K\in\mathbb{R}^{r\times n}$，使得闭环系统具有 $\operatorname{rank}E$ 个有限极点的充分必要条件是广义系统 $\Sigma(E,A,B,C)$ 是脉冲能控的.

3. 如果 $\Sigma(E,A,B,C)$ 是强能控的，则必存在 $K\in\mathbb{R}^{r\times n}$，使得闭环系统 $\Sigma_C^{(1)}(E,A+BK,B,C)$ 有 $\operatorname{rank}E$ 个可任意配置的有限极点，即对任给的 $\operatorname{rank}E$ 个共轭成对的复数集合 Λ_1，总可找到 $K\in\mathbb{R}^{r\times n}$，使得 $\sigma(E,A+BK)=\Lambda_1$.

状态及其导数反馈(state and its derivative feedback) 广义系统的一种反馈形式. 它是含有状态及其导数的反馈. 给定广义系统

$$\Sigma:\quad \begin{cases} E\dot{x} = Ax + Bu, \\ y = Cx. \end{cases}$$

状态及其导数反馈为

$$u = K_1x - K_2\dot{x} + v,$$

其中 $K_1,K_2\in\mathbb{R}^{r\times n}$. 这时闭环系统为

$$\Sigma_C^{(2)}:\quad \begin{cases} (E+BK_2)\dot{x} = (A+BK_1)x + Bv, \\ y = Cx. \end{cases}$$

对任给的 n 个共轭成对的复数集合 Λ_2，存在状态及其导数反馈，使得

$$\sigma(E+BK_2, A+BK_1) = \Lambda_2$$

的充分必要条件是 $\Sigma(E,A,B,C)$ 为能控的. 上述结论说明：如果 $\Sigma(E,A,B,C)$ 是能控的，通过状态及其导数反馈，不但可以任意移动其有限极点，而且亦可以把其 $n-\deg\det(sE-A)$ 个无穷极点都移到复平面中任意有限位置上，实际上通过状态及其导数反馈，可将原能控的广义系统 $\Sigma(E,A,B,C)$ 变成任意一个 n 阶的正常系统. 因此 $\Sigma(E,A,B,C)$ 能控的第一个条件 $\operatorname{rank}(E,B)=n$ 亦称能正常化条件.

广义系统的状态观测器(state observer of singular system) 由广义线性定常系统的输出获得状态估计的一种算法或装置. 对给定的广义系统 $\Sigma(E,A,B,C)$，如果存在如下动态系统

$$\begin{cases} E_C\dot{x}_C = A_Cx_C + B_Cu + Gy, \\ w = F_Cx_C + Fy + Hu, \end{cases} \tag{1}$$

其中 $x_C\in\mathbb{R}^{n_C}$，$w\in\mathbb{R}^n$，E_C,A_C,B_C,G,F_C,F,H 为适当维常数阵，使得

$$\lim_{t\to\infty}(w-x(t))=0\quad(\forall x_0,x_{C0}),$$

则称动态系统(1)为广义系统 $\Sigma(E,A,B,C)$ 的状态观测器. 如果 $\operatorname{rank}E_C<n_C$，则称动态系统(1)为广义系统的广义状态观测器；如果 $\operatorname{rank}E_C=n_C$，则称动态系统(1)为广义系统的正常状态观测器. 显然，如果 $\Sigma(E,A,B,C)$ 是能检测的，则

$$\begin{cases} E\dot{x}_C = (A-GC)x_C + Bu + Gy, \\ w = x_C \end{cases}$$

就是 $\Sigma(E,A,B,C)$ 的广义状态观测器.

广义系统正常状态观测器的存在条件(existence of normal state observer for singular system) 广义系统存在普通观测器的一组条件. 由于广义状态观测器在工程上是不能实现的，因此必须考虑广义系统的正常状态观测器. 关于广义系统正常状态观测器有如下结论：

1. 设广义系统 $\Sigma(E,A,B,C)$ 是能检测的，且 $\operatorname{rank}(E^T,C^T)^T=n$. 如果 $\operatorname{rank}C=m$，则一定存在如下形式的 $n-m$ 阶正常状态观测器：

$$\begin{cases} \dot{x}_C = A_Cx_C + B_Cu + Gy, \\ w = F_Cx_C + Fy. \end{cases}$$

2. 设广义系统 $\Sigma(E,A,B,C)$ 是强能检测的，则一定存在一个如下形式的 $\operatorname{rank}E$ 阶的正常状态观测器：

$$\begin{cases} \dot{x}_C = A_Cx_C + B_Cu + Gy, \\ w = F_Cx_C + Fy + Hu. \end{cases}$$

广义系统的动态补偿器(dynamic compensator of singular system) 广义系统的一种动态控制器. 给定广义系统 $\Sigma(E,A,B,C)$，如果动态系统

$$\begin{cases} E_C\dot{x}_C = A_Cx_C + B_Cu + Gy, \\ u = F_Cx_C + Fy, \end{cases} \tag{1}$$

其中 $x_C\in\mathbb{R}^{n_C}$，E_C,A_C,B_C,F_C,F 为适当维常数阵，使得闭环系统是稳定的，即

$$\sigma\left(\begin{pmatrix} E_1 & 0 \\ 0 & E_C \end{pmatrix}\begin{pmatrix} A+BFC & BF_C \\ B_CC & A_C \end{pmatrix}\right)\subset\mathbb{C}^-,$$

$\mathbb{C}^-$ 表示左半开复平面，则称动态系统(1)为广义系统 $\Sigma(E,A,B,C)$ 的动态补偿器.

如果 $\operatorname{rank}E_C<n_C$，则称(1)为广义动态补偿器；如果 $\operatorname{rank}E_C=n_C$，则称(1)为正常动态补偿器. 利用广义系统 $\Sigma(E,A,B,C)$ 的状态反馈和观测器理论知，广义系统 $\Sigma(E,A,B,C)$ 存在如下广义动态补偿器

$$\begin{cases} E\dot{x}_C = (A-GC)x_C + Bu + Gy, \\ u = Kx_C \end{cases}$$

的充分必要条件是广义系统 $\Sigma(E,A,B,C)$ 既能稳又能检测.

广义系统正常动态补偿器的存在条件(existence of compensator with normal dynamics for

singular system） 广义系统存在普通动态补偿器的一组条件.由于广义动态补偿器工程上难于实现，因此有必要讨论其正常动态补偿器.给定广义系统$\Sigma(E,A,B,C)$，关于其正常动态补偿器

$$\begin{cases}\dot{x}_C=A_C x_C+B_C u,\\ u=F_C x_C+Fy\end{cases}$$

有如下结论：

1. 广义系统$\Sigma(E,A,B,C)$存在正常动态补偿器的充分必要条件是$\Sigma(E,A,B,C)$为既能稳又能检测.

2. 若广义系统$\Sigma(E,A,B,C)$是能稳、能检测的，且$\operatorname{rank}(E,B)=n$，则一定存在$n$阶正常动态补偿器.

3. 若广义系统$\Sigma(E,A,B,C)$是能稳、能检测的，且满足$\operatorname{rank}(E^{T},C^{T})^{T}=n$，$\operatorname{rank}C=m$，则一定存在$n-m$阶正常动态补偿器.

4. 如果广义系统$\Sigma(E,A,B,C)$是强能稳、强能检测的，则一定存在$\operatorname{rank}E$阶正常动态补偿器，使得闭环系统有$n_C+\operatorname{rank}E=2\operatorname{rank}E$个可任意配置的稳定极点.

有限时间完全能控性(complete controllability in finite time) 控制受限下系统的一种结构性质.它反映控制受限下输入对状态变量的制约能力.在控制受限条件下，对于任何状态x_0，总存在某个满足约束条件的允许控制$u(t)$，它能在有限时间内将以x_0为初始状态的系统转移到状态空间中的原点.控制不受限的完全能控性与控制受限下的有限时间能控性不同，前者是后者的必要条件.但是，要使后者成立，对不同的控制限制，还须加上不同的条件.例如，对常系数线性系统控制幅值受限的情况，就尚须满足$\operatorname{Re}\lambda(A)\leqslant 0$；对于控制能量受限，尚须满足$\operatorname{Re}\lambda(A)\leqslant 0$；对于控制燃料受限，尚须满足$\operatorname{Re}\lambda(A)<0$.上述的$\lambda(A)$表示系统矩阵$A$的任一特征值.

有限时间完全能达性(complete reachability in finite time) 控制受限下系统的一种结构性质.它反映控制受限下控制对状态的制约能力.在控制受限条件下，初始状态位于坐标原点，对于任给的终止状态，总存在某个满足约束条件的允许控制$u(t)$，它能在有限时间内将系统从原点转移到预定的终止状态.控制不受限的完全能达性与控制受限下的有限时间完全能达性不同，前者是后者的必要条件.但是，要使后者成立，对于不同的控制限制，还须加上不同的条件.例如对常系数线性系统控制幅值受限的情况，就尚须满足$\operatorname{Re}\lambda(A)\geqslant 0$；对于控制能量受限，尚须满足$\operatorname{Re}\lambda(A)\geqslant 0$；对于控制燃料受限，尚须满足$\operatorname{Re}\lambda(A)\geqslant 0$.上述的$\lambda(A)$表示$A$的任一特征值.

控制受限的完全能稳性(complete stabilizability under constrained control) 控制受限下系统的一种结构性质.在控制受限条件下，对于任何初始状态x_0，总存在某个满足约束条件的允许控制$u(t)$，将其代入状态方程，得相应于初始状态为x_0的解$x(t;x_0,u(t))$，这个解满足如下条件

$$\lim_{t\to+\infty}x(t;x_0,u(t))=0.$$

控制不受限的完全能稳性与控制受限下的完全能稳性不同，前者是后者的必要条件.但是，要使后者成立，对于不同的控制限制，也还须加上不同的条件.

控制受限的能控域(controllable region under constrained control) 控制受限下系统能控状态的全体.是指按如下方式构成的集合：首先定义N拍能控域$Q(N)$，它是由满足如下条件的一切初始状态x_0组成的集合，即对于初始状态x_0，总存在满足某种控制约束条件的某一N拍允许控制序列，它能将x_0在第N拍转移到坐标原点.令

$$Q=\bigcup_{0\leqslant N<+\infty}Q(N),$$

则把Q称为控制满足某种约束条件的能控域.用矩量论方法与阶的估计方法可以得到各种控制约束条件下的能控域的结构性定理.

控制受限的可稳定域(stabilizable region under constrained control) 控制受限下可稳定初始状态的全体.在控制受限下，设x_0是某一初始状态，若存在某个满足约束条件的允许控制序列$[u(i)]_{i=0}^{\infty}$，将其代入离散状态方程

$$x(k+1)=Ax(k)+Bu(k),$$

得解$x(k;x_0,u(\cdot))$，并且这个解满足条件

$$\lim_{k\to\infty}x(k;x_0,u(\cdot))=0,$$

则称初始状态x_0是可稳定的，把一切这样的初始状态所组成的集合称为控制受限(某种约束)下的可稳定域.

离散系统的最小拍控制(minimum step control for discrete system) 离散系统的一类特殊受限控制序列.在控制受限下，能将初始状态用最少步数转移到原点的控制序列.注意，只有当初始状态属于能控域时，最小拍控制序列才存在.此外，在控制幅值受限下，定常离散系统的最小拍控制与定常连续系统的快速控制在性质上不同.对于能控域中的任一初始状态，后者是存在惟一的，有综合方法，而且一定是bang-bang控制；而前者只不惟一存在，不一定是bang-bang控制.目前，对于控制幅值受限与控制燃料受限下，定常离散线性系统的最小拍控制如何求解尚未完全解决.

矩量论方法(moment method) 一种求取离散系统控制序列的方法.它类似于统计学中已知各种矩，求其分布密度函数的方法.在控制理论上可以利用l矩量论方法来求控制受限下的快速控制与最小

拍控制,同样也可以用来解决控制受限下的能控域的结构问题.1938年,苏联数学家克列因(Крейн,M. Г.)首先对l矩量论问题作了详细的讨论.1957年,苏联学者克拉索夫斯基(Красовский,H. H.)利用这个方法求定常连续线性系统的快速控制,其思路如下,由定常线性系统理论可知

$$x_0 = -\int_0^T e^{-At}Bu(t)dt \quad (\|u(t)\| \leqslant M).$$

在上式中,可将x_0看做是已知矩,$u(t)$看做矩分布密度函数,$e^{-At}B$看做权重因子,于是求快速控制$u(t)$的问题,变成已知矩与权重因子,求满足上式的最小周期T与矩分布密度函数的问题.

阶的估计方法(order estimation method) 解决控制受限下的有限时间完全能控性与能控域的结构等问题的一种方法.对于定常离散线性系统

$$x(k+1) = Ax(k) + bu(k) \quad (|u(k)| \leqslant M),$$

它的有限时间完全能控性归结为对于任给的n维单位向量ξ,如下的极限

$$\lim_{N\to\infty}\sum_{i=0}^{N-1}|\xi^T A^{-i-1}b|$$

是否有限.这与A的特征值的结构有关.

2D系统(two-dimensional system or 2D system) 一类特殊的时间空间系统.凡仅在时间和空间各维度中的两个维度上具有相互独立的动态的空间、时间系统均可称为2D系统.早期亦曾译作二维系统,但由于这种译法极易与系统中状态空间的维数,或输入、输出向量的维数及有关物理意义上的维数(如某些分布参数系统,特别是热扩散系统的维数称谓)混淆,后国内有关学者共同约定采用2D系统的术语,相似的情形还有nD系统等.这类系统最明显的特征之一就是其离散情形(或连续情形离散化模型)均可用一组双指标向量序列$\{x_{ij}\}$来描述其运动过程.如一维热扩散系统、二维稳态①常系数分布参数系统、各种2D线性离散状态空间模型、各类二维滤波器二元网络②及二维图象处理系统等均可称为2D系统或纳入2D系统范畴.2D系统③的状态空间的维数可以是有限维,也可以是无穷维的.

注:①这里意为系统方程中不含$\partial u/\partial t$的系统;②源于原文multivariate network一词;③其实,若借用"二元函数"、"多元分析"一类的术语,将2D系统称为二元系统是最恰当不过了,但时下任何文献中并无此称谓.

2D系统理论(2D system theory) 主要指2D控制系统理论.是研究2D系统的建模方法及其各种动态特性,如能控性、能观性、稳定性、可辨识性等分析特性,以及各类控制综合设计问题,如各种最优辨识和滤波理论及各类鲁棒调节器和确定性或随机性反馈控制设计理论等.与1D情形不同,目前2D控制系统理论中较为成熟的领域,就连续系统而言,主要还是2D滤波器理论和部分分布参数系统理论等较为传统的理论.而在离散情形,则主要包括2D线性离散状态空间理论和2D数值滤波器理论等较具特色的理论.总的来说,2D离散系统理论的发展,至少在目前似乎已远远超越了2D连续系统理论,这主要是由于当今科学技术的发展正处在一个计算机科学与技术、工程理论与技术以及应用与纯粹数理科学交互渗透的高峰时期的缘故.

2D状态空间理论(2D state space theory) 研究与2D状态空间有关的一些理论.它研究2D系统的各种状态空间模型的实现,各类能控(达)性、能观(重构)性、稳定性(包括BIBO稳定性和一致渐近稳定性)和系统的反馈控制与综合设计,以及随机最优滤波和系统辨识等的理论.

2D状态空间模型(2D state space model) 2D系统的一种状态空间描述.它是一种通过引入2D状态空间及其空间状态方程的方式来描述2D系统的运动过程的系统模型.其中最著名和最常见的状态空间模型有2D诺叶塞尔模型(RM)、2D阿塔西模型(AM)以及第一类与第二类2D弗那西尼-玛切西尼模型(FM Ⅰ,FM Ⅱ),乃至一般模型(GM)等.以离散情形为例,与1D状态空间相比,2D状态空间在如下两个方面有显著的差异和区别,其一是1D状态向量以单指标向量序列$\{x(k)\}$为特征,系统的初始点为初始状态,而2D状态向量却以双指标向量序列$\{x(i,j)\}$为特征,系统除了初始点外,还有边界点(即边界条件或状态);其二,与1D状态不同,2D状态分局部状态和全局状态两种形态,而且除可分系统外,在一般情形下,2D状态空间模型并不像1D情形那样具有有限的特征模态.

全局状态(global state) 2D系统状态的一种分类刻画.全局的定义与边界条件的具体形式有关.由于受2D状态空间模型的边界条件影响,系统状态无法像1D情形那样简单地对其过去、现在、将来等性质进行刻画,从而产生了局部状态和全局状态的概念.形象地讲,即若设已知边界条件为"第0级状态",则某个能反映处于其后第k级的全体局部状态的集合就可定义为全局状态x_k.以对角边界条件$\{x(i,-i):i=0,\pm1,\pm2,\cdots\}$为例,则全局状态可定义为$x_k=\{x(i,j):i+j=k\}$,而以标准边界:$\{x(i,0),x(0,j):i,j=0,1,2,\cdots\}$为例,则全局状态可更具体地定义为$x_{k,l}=\{x(k+i,l),x(k,l+j):i,j=0,1,2,\cdots\}$.而关于2D全局状态的严格的抽象定义则涉及模论等专业领域,这里所取的是库莱克(Kurek)的描述方法.系统的每一个单一的状态$x(i,j)$相对于全局状态而言,均为局部状态.

局部状态(local state) 见"全局状态".

可分系统(separable system)　在一定意义下可分的2D系统. 它是指其2D传递函数矩阵中各元素的公分母(亦即其二元特征多项式)可分解为两个一元多项式的乘积的线性定常(通常为离散的)系统. 也称为具有可分分母的2D系统. 这类系统的2D特征值可通过两个1D系统的特征值的组合来表示，只要这些特征值落在双圆盘$\{(z_1, z_2): |z_1|\leqslant 1, |z_2|\leqslant 1\}$之内，则系统即为一致渐近稳定的，反之亦然，故与1D情形一样. 这类系统的动态稳定行为确由系统的有限个特征模态所决定.

2D(**离散**)**传递函数矩阵**(2D(discrete)transfer function matrix)　2D系统的一种频域描述. 它是由各种2D线性常系数离散状态空间模型(或系统输入及输出向量)通过二维Z变换得到的系统输入和输出关系$Y(Z_1, Z_2)=G(Z_1, Z_2)U(Z_1, Z_2)$中起传递关系作用的系数矩阵$G(Z_1, Z_2)$.

随机控制系统

随机控制系统(stochastic control system)　受随机因素影响的动态系统. 通常用随机微分方程或随机差分方程来描述. 从控制理论的角度来看，对随机系统的研究主要包括下列方面内容：

建模与辨识. 从量测数据建立系统的数学模型，或已给出系统的模型结构，根据量测数据来估计模型中的未知参数.

滤波. 根据量测数据及系统的模型，估计受到噪声干扰的系统状态或信号.

随机控制. 对随机系统构造控制，使给定的性能指标达到极小.

随机适应控制. 对随机系统一方面辨识系统，估计参数，同时又给出控制，使性能指标达到最小.

具体来讲，在建模与辨识方面，对量测到的随机数据，常用白噪声驱动差分方程(ARMA过程)来建模. 对ARMA过程的系数、阶次等的估计称时间序列分析，它已有一套较成熟的方法，但能用这种方法有效处理的过程通常要有平稳性，而反馈控制系统，由于控制项的作用，系统的输出过程一般不具有平稳性，所以常用的时间序列分析方法对反馈控制系统的参数估计并不适用. 对随机控制系统建模时最常用的模型是ARMAX过程，对它的系数估计通常用最小二乘、极大似然或由此引申出来的其他算法来估计，但要使估计收敛到真值，就得要求系统受到一定程度的激励. 但这种激励要多大，收敛速度有多快，这就引发了许多研究. 对反馈控制系统的阶估计是一个饶有兴趣的问题.

对滤波来讲，经典的维纳-柯尔莫哥洛夫滤波是针对信号和噪声都是平稳过程的情形提出来的. 当信号是动态系统的输出时，它就不再是平稳过程，也就无法用经典的滤波方法. 这就产生了卡尔曼滤波. 当系统的状态方程和量测方程都由高斯变量驱动，并且系统对状态线性，但允许对量测非线性依赖，这时卡尔曼滤波可以写成一组封闭的方程，在实际中得到广泛应用. 但当系统非线性地依赖状态时，除了一些特殊情形外，滤波方程无法解出来，最多做些近似计算. 非线性滤波至今仍是不断引人研究的课题.

在随机控制方面，对部分观测的系统，状态不能精确地量测，最多只能得到它的滤波值，这时一个直观的想法是用状态滤波值来取代相应确定性系统(即把噪声取为零)最优控制中的状态变量，称为必然等价控制. 但这样的控制未必就是最优随机控制. 一个重要的例外是线性二次高斯(LQG)问题，对它的必然等价控制正是最优随机控制. 对ARMAX系统，就不用滤波，对常见的LQ、跟踪、模型参考等问题均可得到最优随机控制. 但一般说，只有线性系统，并只对有限的几种性能指标，才能得到最优随机控制. 对非线性系统，除了极个别例外，不易得到最优随机控制的显式表达. 用次优控制逼近，是一个有效的方法. 随机极大值原理，虽有一定指导意义，但难于应用.

在随机适应控制中，已有的结果主要集中在完全观测的随机控制系统，它要求一边估计系统参数，一边设计随机控制. 如果系统又是部分观测的，那么还要加滤波. 三项任务要同时完成，至今还未见到一个完整的结果. 对ARMAX系统的适应LQ问题、适应镇定、适应模型参考等问题，都已有理论上严格的结果. 特别是适应跟踪，它在工程实际中已有广泛应用，其收敛性和最优性证明也已经解决.

由于工程技术、环境生态、社会经济等领域中出现的实际系统，一般都带有随机干扰，所以对随机系统的研究为实际应用的必须，但也带来许多艰深的理论课题，所以长期以来，随机控制系统受到各种背景研究者的广泛关注.

极大似然估计(maximum likelihood estimation)　参数估计的一种有效方法. 设R^n中一随机向量为$y^n=(y(1), y(2), \cdots, y(n))$，并设其联合概率密度函数为$f(\theta; x_1, x_2, \cdots, x_n)=f(\theta; x^n)$，其中$\theta$为有限维参数向量. 观测值(或实现)确实应该取值$y^n$的概率，与下列值成正比$f(\theta; y_1, y_2, \cdots, y_n)=f(\theta; y^n)$. 上式称为似然函数. 一旦观测值给定，它就是$\theta$的一个确定性函数. 一个合理的想法是选取$\theta$的估计值使得观测到的事件尽可能相像，也就是优化下式

$$\max_{\theta} f(\theta, y^n),$$

使上式达极大值的向量$\hat{\theta}_{\mathrm{ML}}(y^n)$称为$\theta$的极大似然

估计(MLE). MLE 在系统辨识、统计等方面有广泛应用,它具有很好的渐近性质. 比如,假若$\{y(i)\}$是独立同分布序列,那么可以证明$\hat{\theta}_{ML}(y^n)$是强相容的,且$J_n(\hat{\theta}_{ML}(y^n)-\theta)$是渐近正态的,渐近均值为零,方差矩阵与克莱姆-拉奥下界相等.

渐近有效无偏估计(asymptotically efficient unbiased estimation) 一种参数估计方法. 它是对随机向量的联合概率密度函数中未知参数的一种估计. 对 R^n 中的一随机向量 $y^n=(y(1),y(2),\cdots,y(n))$,设它的联合概率密度函数为

$$f(\theta;x_1,x_2,\cdots,x_n)=f(\theta;x^n),$$

其中 θ 为有限维参数向量. 假设 $\hat{\theta}(y^n)$为 θ 的无偏估计,即 $E_\theta\hat{\theta}(y^n)=\theta$,估计误差的方程为

$$V_n(\theta,y^n)=E_\theta(\hat{\theta}(y^n)-\theta)(\hat{\theta}(y^n)-\theta)^\tau,$$

其中 E_θ 是在概率密度 $f(\theta;x^n)$下对 y^n 求数学期望. 则在较弱的正则条件下,有克莱姆-拉奥不等式

$$V_n(\theta,y^n)\geqslant I_n^{-1}(\theta),$$

其中

$$I_n(\theta)=E_\theta\left[\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\lg f(\theta;y^n)\right],$$

称之为费希尔信息矩阵. 进一步地,若极限

$$V(\theta)=\lim_{n\to\infty}V_n(\theta,y^n)$$

存在,并且

$$V(\theta)=\lim_{n\to\infty}I_n(\theta)$$

成立,则称 $\hat{\theta}(y^n)$为 θ 的渐近有效无偏估计.

强一致性(strong consistency) 参数估计的一种收敛性质. 指参数估计值以概率为 1 收敛到参数真值. 又称为强相容性. 设 $y^n=(y(1),y(2),\cdots,y(n))$为 R^n 中一随机向量,$\hat{\theta}(y^n)$为有限维参数 θ 的一个估计,若 $\hat{\theta}(y^n)\to\theta$,当 $n\to\infty$,则 $\hat{\theta}(y^n)$称为一致估计. 若 $\hat{\theta}(y^n)$以概率 1 收敛于 θ,则称估计 $\hat{\theta}(y^n)$具有强一致性.

一致估计(consistent estimation) 见"强一致性".

持续激励(persistence of excitation) 一种为保证系统参数估计一致性对系统输入、输出信号要求的条件. 若矩阵

$$C_n=\begin{bmatrix} C(0) & C(1) & \cdots & C(n-1) \\ C(1) & C(0) & \cdots & C(n-2) \\ \vdots & \vdots & & \vdots \\ C(n-1) & C(n-2) & \cdots & C(0) \end{bmatrix}$$

正定,其中

$$C(k)=\lim_{t\to\infty}\frac{1}{t}\sum_{i=1}^{t}u(i)u(i-k)$$

是平方可求和信号 u 的经验协方差,则称 u 为 n 阶持续激励信号. 对输入信号 u 加上持续激励条件是为了能够用最小二乘方法估计有限脉冲响应模型

$$y(t)=b_1u(t-1)+b_2u(t-2)+\cdots+b_nu(t-n)$$

的参数 $\theta^\tau=[b_1,b_2,\cdots,b_n]$.

阶估计的赤池准则(Akaike criterion for order estimation) 简称阶估计的 AIC 准则. 一种用于估计阶数的准则函数. 在拟合自回归模型时,如果模型的阶数未知,则要对其估计. 估计的方法有多种,其中一类是以优化某种准则函数的解作为阶估计. AIC 准则是应用最普遍的一种,其形式如下:

$$\mathrm{AIC}(k)=\lg\sigma^2(k)+\frac{2k}{n}\quad(k=0,1,\cdots,p),$$

其中 $\sigma^2(k)$是以样本长度为 n 的数据序列拟合 k 阶自回归模型的残差方差估计,p 为比 n 小的适当正整数. 最小化准则 AIC(k),即求 $\hat{k}$ 使得

$$\mathrm{AIC}(\hat{k})=\min\{\mathrm{AIC}(k)\mid k=0,1,\cdots,p\}.$$

取 $\hat{k}$ 作为自回归阶的估计,此即由 AIC 准则给出的阶估计. 此准则由日本学者赤池弘次(Akaike)首先提出,最早用于时间序列分析,现已被广泛地用于有限参数模型的阶数估计,以及变元选择. 理论分析表明:此种估计得到的阶比系统的真阶低的概率渐近为零,而偏高的概率是正的.

阶估计的修正赤池准则(modified Akaike criterion for order estimation) 亦称贝叶斯准则,简称阶估计的 BIC 准则. 一种用于估计未知阶数的准则函数. 在拟合自回归模型时,为估计其阶数,类似于 AIC 准则(参见"阶估计的赤池准则"). BIC 准则是 AIC 准则的一种修正形式,具体形式如下:

$$\mathrm{BIC}(k)=\lg\sigma^2(k)+\frac{k\lg n}{n}\quad(k=0,1,\cdots,p),$$

其中 $\sigma^2(k)$是以样本长度为 n 的数据序列拟合 k 阶自回归模型的残差方差估计,p 为比 n 小的适当正整数. 最小化准则 BIC(k),即求 $\hat{k}$ 使得

$$\mathrm{BIC}(\hat{k})=\min\{\mathrm{BIC}(k)\mid k=0,1,\cdots,p\}.$$

取 $\hat{k}$ 作为自回归阶的估计,此即由 BIC 准则给出的阶估计. 理论分析表明:BIC 准则与贝叶斯方法有关,而且当样本长度 n 增大时,上述的阶估计可任意接近真阶.

阶估计的控制准则(control criterion for order estimation) 简称阶估计的 CIC 准则. 对带输入的自回归滑动平均模型阶估计的一种判别准则. 它与最小二乘法(或其他参数估计方法)相结合,能同时估计出控制系统模型 ARMAX(n,m,r)的阶次 n,m,r 及相应的参数与噪声方差,它的数学表达式如下:

$$\mathrm{CIC}(n,m,r)_t=\hat{\sigma}(n,m,r)_t+(n+m+r)s_t,$$

这里 t 是采样时刻,$\hat{\sigma}(n,m,r)_t$ 是模型噪声方差的估计,s_t 是以一定速度发散到$+\infty$的正实数序列. CIC 准则的具体使用步骤为对不同的阶次 n,m,r,首先用最小二乘法(或其他参数估计方法)估计相应阶次的模型参数,然后计算 $\hat{\sigma}(n,m,r)_t$ 的值及相应的

CIC$(n,m,r)_t$ 的值. 最后选择使 CIC$(n,m,r)_t$ 的值达最小的 n,m,r 作为 t 时刻系统模型阶次的估计. CIC 准则与 AIC 准则、BIC 准则的区别在于它不要求量测数据的平稳性,因而能用于自适应反馈控制系统. 这一准则是中国学者较早提出和研究的.

随机梯度法(stochastic gradient method) 简称 SG 法. 一种参数估计算法. 源自著名的罗宾斯-门罗算法,它有如下递推公式:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \frac{\varphi_t}{r_t}(y_{t+1}^{\mathrm{T}} - \varphi_t^{\mathrm{T}}\hat{\theta}_t),$$

$$r_t = 1 + \sum_{i=0}^{t} \|\varphi_i\|^2,$$

这里 t 是采样时刻,y_t 是实测数据,φ_t 是回归量的实测数据,X^{T} 表示矩阵 X 的转置,$\hat{\theta}_t$ 是未知参数在 t 时刻的估计值. 较之最小二乘法,随机梯度法的优点是计算简单,缺点是收敛速度慢.

常微分方程方法(ordinary differential equation method) 研究递推算法收敛性的一种方法. 考察递推算法 $x_{k+1}=x_k+a_k(f(x_k)+\varepsilon_{k+1})$, $a_k>0$ 是算法步长,ε_{k+1}是误差项,通常

$$a_k \underset{k\to\infty}{\longrightarrow} 0, \quad \sum_{k=1}^{\infty} a_k = \infty.$$

把$\{x_k\}$以$\{a_k\}$为步长做线性内插函数 x_t,考察尾函数 $x_n(t) \triangleq x_{t+n}$所组成的函数组$\{x_n(t)\}$. 在一定的条件下,$\{x_n(t)\}$一致有界且等度连续,那么存在子序列$\{x_{n_k}(t)\}$在有穷区间内一致收敛到一个连续函数 $x(t)$,它满足微分方程

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f(x(t)).$$

常微分方程方法证明,算法的收敛性和上述微分方程的稳定性有密切联系. 粗略地说,在一些条件下,当 x^0 是微分方程的稳定解时,则 $x_n \to x^0$.

高斯-马尔可夫估计(Gauss-Markov estimation) 一种线性回归模型参数的估计方法. 以 y_t 表示因变量,$x_{1t},x_{2t},\cdots,x_{pt}$表示 p 个自变量,它们满足如下的线性回归模型:

$$y_t=\beta_1 x_{1t}+\cdots+\beta_p x_{pt}+\varepsilon_t \quad (t=1,2,\cdots,n). \tag{1}$$

若记

$$\underline{\beta}=(\beta_1,\beta_2,\cdots,\beta_p)^{\mathrm{T}}, \ \underline{\varepsilon}=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)^{\mathrm{T}},$$
$$X=(\underline{x}_1,\underline{x}_2,\cdots,\underline{x}_p), \ \underline{x}_i=(x_{i1},x_{i2},\cdots,x_{in})^{\mathrm{T}},$$
$$\underline{Y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}},$$

则(1)式写成

$$\underline{Y}= X\underline{\beta} + \underline{\varepsilon}. \tag{2}$$

在模型(2)中,若假定 $\underline{\varepsilon}$服从 n 维正态 $N(0,\sigma^2 I_n)$分布,其中 I_n 为 n 阶单位阵,(2)式被称为高斯-马尔可夫模型. 回归参数 β 的高斯-马尔可夫估计,与通常的最小二乘估计相同,即

$$\hat{\underline{\beta}}= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\underline{Y}. \tag{3}$$

若假定 $\underline{\varepsilon}$服从 n 维正态 $N(0,R)$分布时,(2)式中的回归参数的高斯-马尔可夫估计为

$$\hat{\underline{\beta}}= (X^{\mathrm{T}}R^{-1}X)^{-1}X^{\mathrm{T}}R^{-1}\underline{Y}. \tag{4}$$

它不同于(3)式的最小二乘估计,当 X 为常数矩阵时,(4)式的估计是线性的无偏的最小方差估计.

近似极大似然递推估计(approximate maximum likelihood recursive estimation) 简称 AML 递推估计. 一种参数估计的方法. 全称为近似极大似然递推估计. 极大似然估计是选择参数 θ,使得在已知量测数据 Y 时似然函数 $P(Y|\theta)$极大,这里 $P(Y|\theta)$是数据 Y 的概率分布函数. 与最小二乘估计不同的是极大似然估计需要已知似然函数,实际上这是很困难的. 通常可假设 $P(Y|\theta)$是正态分布函数,这时极大似然估计与最小二乘估计相同. 为了减少计算量,产生了近似极大似然递推估计,把求似然函数极大问题精化为残差平方和的极小问题. 进一步把残差平方和近似为参数 θ 的二次函数,就可以得到参数 θ 的递推估计,即已知时刻 t 上的参数估计、量测数据及时刻 $t+1$ 上的量测数据,计算时刻 $t+1$ 上的参数估计. 已经证明在一定条件下 AML 递推估计以概率 1 收敛到估计准则的一个局部极小值.

激励条件(excitation condition) 一种加于观测数据的条件. 指在系统辨识中,为了使参数估计相容而对观测数据所加的条件. 特别地,对线性回归模型,若 φ_i 为回归向量,则激励条件可转化为对矩阵 $\sum_i \varphi_i\varphi_i^{\mathrm{T}}$ 的要求.

系统辨识(system identification) 一种统计建模的理论和方法. 系统辨识的主要内容是模型结构选择、试验设计、参数估计和模型验证. 而模型结构选择和参数估计是其中的基本问题,其主要目的是辨识系统的未知量. 在控制问题中,试验设计简称输入信号的选择. 在开环条件下,输入信号的选择比较简单,但在自适应控制中,当加到对象的输入信号由反馈产生时,就会遇到困难. 在有些情况下,不能惟一地确定出待估计的参数,此时可能需要引入摄动信号才能解决这个问题. 在自适应控制中,过程参数在不断地变化着,所以应有递推地更新参数的估计方法,即递推估计方法. 常用的递推估计方法有递推最小二乘法、递推极大似然法、递推辅助变量法等. 递推估计的当前估计值只依赖于前面的估计和当前量测. 模型结构选择的主要内容为模型的阶次、时滞、模型种类等模型结构参数的选择. 在求解辨识问题时,正确验证所得结果也是十分重要的.

递推估计方法(recursive estimation method)见"系统辨识".

可辨识性(identifiability) 系统的一种基本属性. 指系统的模型可以通过系统的输入、输出量测数

据惟一确定的性质. 可辨识性是系统辨识的根本性质,有两种可辨识性:一种先验可辨识性,是由系统及输入、输出的结构决定的,而不受数据的误差影响;另一种后验可辨识性,是受到量测数据有误差或数目少的影响,造成参数估计有误差而出现不可辨识问题,这时就要考虑极限情形,即当数据不断增加时,参数估计值能否收敛到惟一的参数值. 在出现不可辨识现象时,有时能够利用改变系统及输入、输出的结构或者重新设计实验,减少量测数据或者改变辨识的方法以便达到可辨识性. 例如,闭环条件下量测的输入、输出数据辨识开环系统的模型,有时会发生不可辨识问题,通过增加反馈通道的阶次或叠加白噪声信号到量测数据上,会改善可辨识性.

自回归滑动平均系统(autoregressive-moving average system) 简称 ARMA 系统. 一类随机系统. 它是由线性定常方程

$$\sum_{i=0}^{p}A(i)Y(t-i)=\sum_{i=0}^{q}B(i)\varepsilon(t-i)$$

表示的系统,其中 $Y(t),\varepsilon(t)\in\mathbb{R}^m$,分别为系统的输出和干扰向量,$A(i),B(i)\in\mathbb{R}^{m\times m}$,$t$ 为离散时间参数,$\{\varepsilon(i)\}$为白噪声序列,$\{Y(i)\}$为上述差分方程的解,p,q 分别为系统输出和噪声的阶数. 上述ARMA系统常记为 ARMA(p,q). 特别地,在ARMA(p,q)中,若 $q=0$,则称之为 AR(p)模型;若 $p=0$,则称之为 MA(q)模型.

带外源变量的自回归滑动平均系统(autoregressive-moving average with exogenous variables system) 简称 ARMAX 系统. 一类随机系统. 它是由下式描述的线性定常随机系统:

$$\sum_{i=0}^{p}A(i)Y(t-i)=\sum_{i=0}^{q}B(i)u(t-i)+\sum_{i=0}^{r}C(i)\varepsilon(t-i),$$

其中 $y(t),u(t)$和 $\varepsilon(t)$分别表示系统的 m,p 和 m 维输出、输入和干扰向量序列,$A(i),B(j),C(k)(0\leqslant i\leqslant p,0\leqslant j\leqslant q,0\leqslant k\leqslant r)$分别表示相应维数的实矩阵,$t$ 是离散时间参数,可能在一无穷集合上取值,而 p,q 和 r 分别表示系统输出、输入和噪声部分的固定阶数. 显然在 ARMAX 的递推中,初值是必要的,若时间 t 从 0 开始,那么就需要给定初值 $\{Y(-1),\cdots,Y(-p),u(-1),\cdots,u(-q),\varepsilon(-1),\cdots,\varepsilon(-r)\}$.

若输入项 $u(t)\equiv0$,那么 ARMAX 系统就成为常见的 ARMA 系统,这个系统在时间序列领域有专门的、深入的研究. 进一步,在 ARMAX 系统中,若 $r=0$,就称之为 ARX 系统;若 $u(t)\equiv0,r=0$,则称之为 AR 模型;若 $u(t)\equiv0,p=0$ 时,则称之为 MA 模型. 对于线性随机系统来讲,其 ARMAX 描述、状态空间描述和传递函数描述是相互等价的. 然而,当考虑辨识、预报等问题时,ARMAX 模型却较为方便. 需要指出的是,在系统和控制理论中对输入信号$u(t)$的要求与时间序列中所考虑的有很大不同. 这里的输入信号是控制输入,它一般是系统的过去输出和输入值的函数,因此系统的输出和输入序列一般是非平稳的.

最小二乘估计(least-squares estimate) 估计数学模型中参数的最古老而又最熟知的方法. 它的最早应用者可以追溯到 17 世纪末的高斯(Gauss, C. F.). 关于这个方法的各种研究文献浩瀚,应用领域极其广泛. 在此只着重介绍在控制系统中的应用形式及递推算法. 考虑下列回归模型:

$$y_{k+1}=\theta_0^{\mathrm{T}}\varphi_k+w_{k+1},\tag{1}$$

其中 θ_0,φ和 w 分别为未知参数、回归向量和模型误差. 在时刻 N,参数 θ_0 的最小二乘估计由下式给出:

$$\theta_N=\arg\min_{\theta}\frac{1}{N}\sum_{i=1}^{N}\|y_i-\theta^{\mathrm{T}}\varphi_{i-1}\|^2.$$

其显式解可表示为

$$\theta_N=\Big(\sum_{i=1}^{N}\varphi_i\varphi_i^{\mathrm{T}}\Big)^{-1}\sum_{i=1}^{N}\varphi_iy_{i+1}.$$

为简单起见,其中假定了矩阵 $\big(\sum_{i=1}^{N}\varphi_i\varphi_i^{\mathrm{T}}\big)$ 的可逆性. 为了计算方便,常将估计写成下列递推形式:

$$\theta_{n+1}=\theta_n+a_nP_n\varphi_n(y_{n+1}-\varphi_n^{\mathrm{T}}\theta_n),\tag{2}$$

$$P_{n+1}=P_n-a_nP_n\varphi_n\varphi_n^{\mathrm{T}}P_n,\tag{3}$$

$$a_n=(1+\varphi_n^{\mathrm{T}}P_n\varphi_n)^{-1},$$

其中初值 $P_0\geqslant0,\varphi_0$ 可任取. 关于最小二乘估计,人们最关心的问题之一是 θ_n 的强相容性问题,为理论上的需要,常假定模型误差 w_i 是关于某非降 σ 代数 $\mathscr{F}_i$ 可测的,且

$$E[w_i|\mathscr{F}_{i-1}]=0,\ \sup_i E\|w_i\|^2<\infty.\tag{4}$$

如果$\{\varphi_i\}$是确定性序列,那么 $\theta_n\to\theta_0$ 几乎处处成立的条件为

$$\lambda_{\min}\Big(\sum_{i=1}^{n}\varphi_i\varphi_i^{\mathrm{T}}\Big)\xrightarrow[n\to\infty]{}\infty\ \text{几乎处处成立},$$

其中 $\lambda_{\min}(\cdot)$表示矩阵的最小本征值. 在控制系统的应用中,φ_i 一般不是确定性的,而是可测的随机序列. 黎子良和魏庆荣(1982)证明:若将(4)中第二部分改为

$$\sup_i E[\|w_i\|^{2+\varepsilon}|\mathscr{F}_{i-1}]<\infty\quad(\text{几乎处处成立},\varepsilon>0),\tag{5}$$

那么在条件

$$\lambda_{\min}\Big(\sum_{i=1}^{n}\varphi_i\varphi_i^{\mathrm{T}}\Big)\to\infty,\quad\frac{\lg\lambda_{\max}\Big(\sum_{i=1}^{n}\varphi_i\varphi_i^{\mathrm{T}}\Big)}{\lambda_{\min}\Big(\sum_{i=1}^{n}\varphi_i\varphi_i^{\mathrm{T}}\Big)}\to0\quad(\text{几乎处处成立})\tag{6}$$

下，仍有 $\theta_n \to \theta_0$ 几乎处处成立. 进一步，若考虑更一般的线性回归模型

$$y_{k+1} = \theta_0^{\mathrm{T}} \varphi_k + w_{k+1} + c_1 w_k + \cdots + c_r w_{k-r+1}, \tag{7}$$

那么未知参数除了 θ_0 外还有 $c_i (1 \leqslant i \leqslant r)$. 记 $\theta = [\theta_0^{\mathrm{T}}; c_1, \cdots, c_r]^{\mathrm{T}}$. 对 θ 的估计可应用推广递推最小二乘方法. 这个算法的形式与(2)和(3)相同，只是将回归向量 φ_i 换成向量 $\psi_i = [\varphi_i^{\mathrm{T}}; e_i, e_{i-1}, \cdots, e_{i-r+1}]^{\mathrm{T}}$，这里 e_i 或取为先验误差 $e_i = y_i - \theta_{i-1}^{\mathrm{T}} \psi_{i-1}$，或取为后验误差 $e_i = y_i - \theta_i^{\mathrm{T}} \psi_{i-1}$. 在应用中取后一种更自然些，同时也给理论分析带来方便. 下面只考虑 e_i 为后验误差的情形. 现在除了假定条件(5)，(6)外，若再假定下列严格正实条件：

$$\max_{|z|=1} |c_1 + c_2 z + \cdots + c_r z^{r-1}| < 1,$$

那么仍可证明 $\theta_n \to \theta_0$. 这些结果在控制系统，尤其是自适应控制系统中有重要应用.

预报误差方法(prediction error method) 基于预报误差的参数辨识方法. 设在时刻 $t-1$ 得到了系统的输入输出观测值 z^{t-1}，可以构造 t 时刻系统输出 y_t 的预测如下：

$$\hat{y}(t|\theta) \triangleq g_M(\theta; t, z^{t-1}),$$

其中 $g_M(\theta; \cdot, \cdot)$ 对固定的 θ 是从 $\mathrm{R}_+ \times \mathrm{R}^{tq}$ 到 R^p 的实函数，θ 是有限维参数，$\theta \in M_d \subset \mathrm{R}^d$，函数 $g_M(\theta; t; z^{t-1})$ 给出了系统的一个具体模型 $M(\theta)(\theta \in M_d)$. 模型有效性的一个自然度量是预报误差 $\varepsilon(t, \theta) = y(t) - \hat{y}(t|\theta)$. 这个误差在给定数据 z^{t-1} 后是可以计算的，但由于 $\varepsilon(t, \theta)$ 是 p 维向量，所以通常引入一个标量 $l(t, \theta, \varepsilon(t, \theta))$ 来度量预报误差的大小，这里 $l(\cdot, \cdot, \cdot)$ 是从 $\mathrm{R} \times \mathrm{R}^d \times \mathrm{R}^p$ 到 R_+ 的实函数. 函数 l 的一个典型的取法是

$$l(t, \theta, \varepsilon) = \frac{1}{2} \varepsilon^{\mathrm{T}} \Lambda^{-1} \varepsilon,$$

其中 Λ 是正定矩阵. 现在假设有 N 个观测数据，判定模型 $M(\theta)$ 可用性的一个自然的判据是

$$V_N(\theta, z^N) = \frac{1}{N} \sum_{t=1}^{N} l(t, \theta, \varepsilon(t, \theta)).$$

这样在 $\theta \in M_d$ 中极小化 $V_N(\theta, z^N)$ 可以得到 θ 的估计值 $\hat{\theta}_N$. 常见的许多离线辨识算法都可由此而得出(对应于不同的 $l(t, \theta, \varepsilon)$ 和 $M(\theta)$). 这些方法称为预报误差辨识方法.

随机逼近(stochastic approximation) 一种参数的估计方法. 它是在有随机误差干扰的情况时，用逐步逼近的方式估计参数的方法. 典型的随机逼近问题可描述为：令 $\{\varepsilon_t\}$ 是同分布的随机序列，已知 ε_t 和 x 的某一函数 $Q(x, \varepsilon_t)$，欲求

$$EQ(x, \varepsilon_t) = f(x) = 0$$

的解，式中 E 表示对 ε_t 求数学期望，但并不知道 $\{\varepsilon_t\}$ 的分布，函数 $Q(x, \varepsilon_t)$ 的确切形式也可能是未知的. 然而，对任何选定的 x 值，都能得到(或观测到) $Q(x, \varepsilon_t)$ 的值. 例如，观测值为 $y_t = h(x) + \varepsilon_t$，要求从观测值估计出参数 x，则取随机逼近算法

$$x_{t+1} = x_t - a_t y_{t+1},$$

式中数列

$$a_t > 0, \quad \sum_{t=0}^{\infty} a_t = \infty, \quad \sum_{t=0}^{\infty} a_t^2 < \infty.$$

1951 年，罗宾斯(Robbins, H.)和门罗(Monro, S.)对 $h(\cdot)$ 加了一定条件后证明

$$E|x_t - \theta|^2 \underset{t \to \infty}{\longrightarrow} 0.$$

这是随机逼近的奠基工作. 在有的问题中，要找到的不是 $h(x)$ 的零点，而是 $h(\tilde{x})$ 的极值点 $\tilde{x}$，它满足 $h'(x) = 0$. 这时观测到的仍是 $y = h(x) + \varepsilon$，而不是 $h'(x) + \varepsilon$，故上述算法已不能用于逼近 $\tilde{x}$. 基弗(Kiefer, J. C.)和维尔弗维茨(Wolfowitz, J.)给出求极值的逼近算法

$$\tilde{x}_{t+1} = \tilde{x}_t - a_t \frac{\mathrm{d}y}{\mathrm{d}t} \Big|_{\tilde{x}_t}.$$

数列 a_t 满足上述同样条件，则可逼近极值点 $\tilde{x}$. 从罗宾斯、门罗以来，随机逼近的研究进一步取得了很大进展. 讨论了观测误差不独立的情况和带约束条件的情况，以及 $h(x)$ 具有更一般性质的情形，并且把概率的方法结合使用了微分方程等工具. 随机逼近在最优化问题、系统辨识、适应控制器等方面都有应用.

辅助变量法(auxiliary variable method) 亦称工具变量法. 一种改进的最小二乘参数估计方法. 即设

$$y(t) = \theta_0^{\mathrm{T}} \varphi(t) + v(t),$$

其中 $y(t)$ 是量测的，$\varphi(t)$ 是回归向量，θ_0 是待估参数向量(T 表示转置)，$v(t)$ 是相关噪声，则最小二乘估计

$$\hat{\theta}_{\mathrm{L,S}}(N) = \Big[\sum_{t=1}^{N} \varphi(t) \varphi^{\mathrm{T}}(t)\Big]^{-1} \cdot \sum_{t=1}^{N} \varphi^{\mathrm{T}}(t) y(t)$$

收敛不到真值 θ. 这种情形下，可以用一个辅助向量 $z(t)$，只要：

1. $z(t)$ 与 $v(t)$ 互不相关，即 $Ez(t)v(t) \neq 0$.
2. 矩阵 $\lim\limits_{N \to \infty} \frac{1}{N} \sum\limits_{t=1}^{N} z(t) \varphi^{\mathrm{T}}(t)$ 可逆.
3. $v(t)$ 的均值为零.

则可以证明，当 $N \to \infty$ 时，辅助变量估计量

$$\hat{\theta}_{\mathrm{AV}}(N) = \Big[\sum_{t=1}^{N} z(t) \varphi^{\mathrm{T}}(t)\Big]^{-1} \cdot \sum_{t=1}^{N} z(t) y(t)$$

将趋于(即收敛于)真值 θ_0.

辅助变量法的关键在于如何构造一个满足上述条件的辅助变量，也就是说辅助变量应当与噪声 $v(t)$ 不相关，而与 $\varphi(t)$ 强相关. 因此常见的一种选择

是由输入激励一个确定性的系统而产生的输出和输入组成，另一种是全都取延迟输入作为辅助变量。辅助变量法的基本思想是很可贵的，适当选择辅助变量，可以沟通与其他辨识方法的联系，因此在许多场合都可以用到它。当然还可有递推辅助变量法，它也得到广泛的使用。

工具变量法(instrumental variable method) 即“辅助变量法”。

算法的弱收敛(weak convergence of algorithm) 算法的一种收敛性定义。设对参数 θ 用算法 $\mathscr{A}$ 对数据 y^t 进行运算，如果

$$\mathscr{A}(y^t)=\hat{\theta}_t \xrightarrow{\text{w.p.1}} \min V(\theta),$$

即 $\hat{\theta}_t$ 以概率 1 收敛于 θ 的某一标量函数(例如，误差的平方和)的极小值，也就是 $\hat{\theta}_t(t\to\infty)$ 是弱收敛的，则称算法 $\mathscr{A}$ 是弱收敛的。其意义是指对所有的样本数据，使用算法 $\mathscr{A}$，几乎所有的(除了一个零测集的样本数据外)$\hat{\theta}_t$ 收敛于 $V(\theta)$ 的极小值处(注：$V(\theta)$ 的极小值决定于模型的特性，有些模型存在多个极小值，则此时收敛只能说收敛到 $V(\theta)$ 的局部极小值)。算法 $\mathscr{A}$ 实质上是从数据 y^t 到参数估计 $\hat{\theta}_t$ 的映射，一般说来是相当复杂的。为了解决算法的收敛性，目前主要有两种方法：

1. 把算法 $\mathscr{A}$ 和与之相应的一个确定性微分方程相联系，这样把该方程的稳定性与 $\mathscr{A}$ 的收敛性相对应。这种分析方法是由雍(Ljung, L.)(1974, 1976)提出的。

2. 引入一个函数，该函数起着对应算法 $\mathscr{A}$ 的随机李亚普诺夫函数的作用，使用鞅论研究该函数的收敛性，从而也就解决了算法 $\mathscr{A}$ 的收敛性。这个方法首先为莫尔(Moore, J. B.)和莱德维奇(Ledwich, G. (1980))，以后为索曼(Solo, V. (1978, 1979))所使用。

适应滤波(adaptive filtering) 利用量测信息，逐步改进滤波器参数的过程。已知两个信号 x 和 y，一个滤波器 F，以使信号 y 和 $\hat{y}=Fx$ 两者尽可能接近。这个问题可用图 1 来表示。

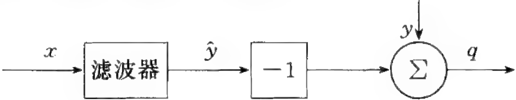


图 1

在典型情况下，有

$$x(t)=s(t)+n(t),\quad y(t)=s(t+\tau),$$

式中，s 是所关心的信号，n 是某个不希望的扰动。当 $\tau<0$ 时，则此问题称为平滑；当 $\tau=0$ 时，则称为滤波；$\tau>0$ 时，则称为预测。当信号性质未知时，相应的问题就成为自适应信号处理问题，而所有的递推参数估计方法都可用于这类问题。图 2 为一个典型的适应滤波问题，其中 $y(t)=s(t)$。调整机构可以是任何一种递推参数估计器。

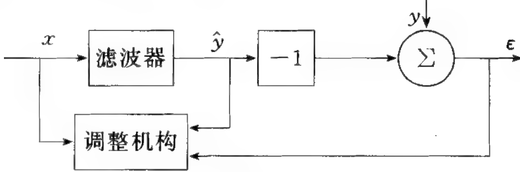


图 2

滤波(filtering) 见“适应滤波”。

平滑(smooth) 见“适应滤波”。

残差(residual) 参数估计中的一种误差。是指由下式

$$\varepsilon(i)=y(i)-\hat{y}(i)=y(i)-\varphi^{\mathrm{T}}(i)\hat{\theta}$$

中表示的量 $\varepsilon(i)$。对于数学模型

$$\begin{aligned} & y(t) \\ &= \varphi_1(t)\theta_1+\varphi_2(t)\theta_2+\cdots+\varphi_n(t)\theta_n \\ &= \varphi^{\tau}(t)\theta, \end{aligned}$$

其中 y 是观测变量，$\theta_1,\theta_2,\cdots,\theta_n$ 为未知参数，$\varphi_1,\varphi_2,\cdots,\varphi_n$ 是由其他已知变量决定的已知函数，t 通常表示时间，记

$$\varphi^{\mathrm{T}}(t)=[\varphi_1(t),\varphi_2(t),\cdots,\varphi_n(t)],$$
$$\theta^{\mathrm{T}}=[\theta_1,\theta_2,\cdots,\theta_n].$$

应用某种估计方法得到 θ 的估计 $\hat{\theta}$，$\hat{y}(i)=\varphi^{\mathrm{T}}(i)\hat{\theta}$ 为估计的观测值，则 $y(i)-\hat{y}(i)$ 称为残差。

谱密度(spectral density) 刻画平稳随机过程特性的一种确定性的量。若 $x(t)$ 是一个平稳随机过程，有自相关函数 $R_x(\tau)=Ex(t+\tau)x^{\mathrm{T}}(t)$，则 $x(t)$ 的谱密度 $\Phi_x(\omega)$ 定义为

$$\Phi_x(s)=\int_{-\infty}^{\infty}R_x(t)\mathrm{e}^{-st}\mathrm{d}t$$

在 $s=\mathrm{j}\omega$ 处的赋值。对于离散时情形，有

$$R_x(k)=Ex_{i+k}x_i^{\mathrm{T}}.$$

谱密度为

$$\Phi_x(z)=\sum_{k=-\infty}^{\infty}R_x(k)z^{-k}$$

在 $z=\mathrm{e}^{\mathrm{j}\omega}$ 处的赋值。称

$$F(\lambda)=\int_{-\pi}^{\lambda}\Phi_x(\omega)\mathrm{d}\omega\quad(-\pi\leqslant\lambda\leqslant\pi)$$

为谱分布函数。

谱分布函数(spectral distribution function) 见“谱密度”。

谱表示定理(spectral representation theorem) 平稳随机过程相关函数的频域表达。若 $u(t)$ 和 $y(t)$ 是脉冲响应为 $h(t)$、传递函数为 $H(s)$ 的一个线性系统的输入和输出，则自相关函数和谱密度有以下关系：

$$R_{yu}(t)=h(t)*R_u(t), \tag{1}$$
$$R_y(t)=h(t)*R_u(t)*h^{\mathrm{T}}(-t), \tag{2}$$

$$\Phi_{yu}(s)=H(s)\Phi_u(s), \quad (3)$$

$$\Phi_y(s)=H(s)\Phi_u(s)H^{\mathrm{T}}(-s), \quad (4)$$

其中 * 表示卷积，$\Phi_{yu}, \Phi_y, \Phi_u$ 为谱密度，则有如下的谱分解定理：若对几乎每个 s，$|\Phi_y(s)|\neq 0$，则 $\Phi_y(s)$ 可惟一地分解成

$$\Phi_y(s) = H(s)\Omega H^{\mathrm{T}}(-s), \quad (5)$$

这里 $H(s)$ 是一个正的实有理传递函数，其所有的极点均在左半开复平面；

$$\lim_{s\to\infty} H(s) = I;$$

$H^{-1}(s)$ 所有的极点均在左半开复平面或在 $\mathrm{j}\omega$ 轴上；$\Omega=\Omega^{\mathrm{r}}>0$ 是实矩阵. 若对所有满足 $\mathrm{Re}(s)=0$ 的 s 还有 $|\Phi_y(s)|\neq 0$，则 $H^{-1}(s)$ 的所有极点均在左半开复平面. 根据(4)和谱分解定理，当输入是谱密度为 Ω 的白噪声时，过程 $y(t)$ 可看成是稳定系统 $H(s)$ 的输出，这就是谱表示定理. 谱密度、谱分解定理等可用于线性随机系统.

谱分解定理(spectral decomposition theorem)　见“谱表示定理”.

沃尔德分解(Wold decomposition)　宽平稳随机过程的一种分解. 设 x 是一宽平稳随机过程，取值于 $\mathbf{R}^p$，且设 e 是其新息过程. 记

$$H_n = \mathrm{Span}\{\cdots, x_{n-1}, x_n\},$$

$$H_{-\infty} = \bigcap_{n\in Z} H_n (x \text{ 的无穷远过去}),$$

$$N_n = \overline{\mathrm{Span}}\{\cdots, e_{n-1}, e_n\}.$$

沃尔德分解定理指出下列事实：

1. $x_n=u_n+v_n$，其中

$$u_n = (x_n|N_n), \ v_n = (x_n|H_{-\infty}),$$

且 $u_n\perp v_n (\forall n\in \mathbf{Z})$.

2. 过程 u 是宽平稳的，且满足

$$u_n = \sum_{k=0}^{\infty} A_k e_{n-k} \quad (n\in \mathbf{Z}),$$

其中 e 是过程 x 的新息，满足

$$Ee_k e_j^{\mathrm{T}} = \sum \delta_{kj} \quad (\forall j,k \in \mathbf{Z}).$$

进一步

$$0 \leqslant \sum_{k=0}^{\infty} A_k A_k^{\mathrm{T}} < \infty,$$

$$Ex_0 e_{-k}^{\mathrm{T}} = Eu_0 e_{-k}^{\mathrm{T}} = A_k \Sigma \ (k\geqslant 0),$$

$$A_0\Sigma = \Sigma = \Sigma A_0^{\mathrm{T}}.$$

若 x 满秩，则序列 $\{A_k | k\in \mathbf{Z}_+\}$ 是惟一的，且 $A_0=I$.

3. v 是线性确定性宽平稳过程(即 $v_n\in \mathrm{Span}\{\cdots, v_{n-2}, v_{n-1}\}, \forall n\in \mathbf{Z}$)，且

$$\overline{\mathrm{Span}}\{\cdots, v_{n-1}, v_n\} = H_{-\infty} \quad (\forall n \in \mathbf{Z}).$$

上述事实明显说明 x 可表示成下列的沃尔德分解：

$$x_n = \sum_{k=0}^{\infty} A_k e_{n-k} + v_n \quad (\forall n \in \mathbf{Z}),$$

其中 e 是一宽平稳正交过程，且与确定性过程 v 正交. 若 x 是纯(线性)非确定性的，即若 $H_{-\infty}=\{0\}$，则 $v_n=0 (\forall n\in \mathbf{Z})$，则有

$$x_n = \sum_{k=0}^{\infty} A_k e_{n-k}.$$

显然，对具有二阶矩的非平稳过程 x，下述沃尔德分解同样也成立：

$$x_n = u_n + v_n = \sum_{i=0}^{\infty} A_{n_i} e_{n-i} + v_n,$$

其中 $A_{n_i} e_{n-i} = Ex_n e_{n-i}^{\mathrm{T}} = Eu_n e_{n-i}^{\mathrm{T}}$，$\Sigma_n = Ee_n e_n^{\mathrm{T}} (\forall n, i\in \mathbf{Z})$，$e_i\perp e_j, i\neq j, u_n\perp v_m \in H_{-\infty} (\forall m, n\in \mathbf{Z})$.

设 x 是满秩宽平稳过程，$Ee_n e_n^{\mathrm{T}}=\Sigma$，用 $\Sigma_c^{1/2}$ 表示使 $(\Sigma_c^{1/2})(\Sigma_c^{-1/2})^{\mathrm{T}}=\Sigma$ 的(惟一的)上三角矩阵，其中 $\Sigma_c^{1/2}$ 的对角线元素全为正的(称为 Σ 的乔莱斯基因子). 记 $\varepsilon_k=\Sigma_c^{1/2} e_k$，$\overline{A}_k=A_k\Sigma_c^{1/2}$，则 $E\varepsilon_k\varepsilon_k^{\mathrm{T}}=I$. 称

$$x_n = \sum_{k=0}^{\infty} \overline{A}_k \varepsilon_{n-k} + v_n$$

为 x 的乔莱斯基正则化沃尔德分解.

维纳-霍普夫方程(Wiener-Hopf equation)　一类特殊的积分方程. 它是使连续时间平稳过程的方差估计误差达到极小的最优滤波器脉冲响应函数满足的必要条件. 设连续时间宽平稳过程 y 和 x 是联合宽平稳的，设基于 $y(s) (-\infty<s<\infty)$ 对 $x(t+\eta) (-\infty<t<\infty)$ 进行估计的线性滤波器由下列卷积给出：

$$\hat{x}(t+\eta) = \int_{-\infty}^{\infty} h(\tau) y(t-\tau)\mathrm{d}\tau \quad (\forall t\in \mathbf{R}),$$

这里 η 未必是正的，所以上述公式可以表述平滑、滤波和预报问题. 根据经典的变分原理，可以求出使均方估计误差 $E(x(t+\eta)-\hat{x}(t+\eta))^2$ 达极小的最优滤波器脉冲响应 h 所满足的必要条件

$$R_{xy}(t+\eta) = \int_{-\infty}^{\infty} h(\mu) R_y(\tau-\mu)\mathrm{d}\mu \quad (\forall \tau\in \mathbf{R}),$$

其中 R_{xy} 表示 x 和 y 的协方差函数，R_y 表示 y 的方差函数. 特别地，若考虑非可料滤波器($h(\mu)=0, \forall \mu<0$)，那么 h 满足

$$R_{xy}(\tau+\eta) = \int_{0}^{\infty} h(\mu) R_y(\tau-\mu)\mathrm{d}\mu \quad (\forall \tau\in \mathbf{R}).$$

上述两式均称之为维纳-霍普夫方程.

线性新息过程(linear innovation process)　随机过程中的信息增加量的反映. 设 $x_n (-\infty<n<\infty)$ 是定义于 $(\Omega, \mathscr{F}, P)$ 上的实随机过程，用 H_n 表示 $\{\cdots, x_{n-1}, x_n\}$ 张成的线性空间 $\mathrm{Span}\{\cdots, x_{n-1}, x_n\}$，$(x_n|H_{n-1})$ 表示 x_n 到 H_{n-1} 的正交投影. 此处假定 $Ex_n^2<\infty (\forall n)$，或 $x_n\in L^2(\Omega, \mathscr{F}, P)$，内积 $(x,y)=Ex^{\mathrm{T}}y (x, y\in L^2(\Omega, \mathscr{F}, P))$. 定义

$$e_n = x_n - (x_n|H_{n-1}) \quad (n\in \mathbf{Z}, \mathbf{Z} \text{ 是正整数集}),$$

那么 $\{e_n\}$ 称为过程 x 的(线性)新息过程. 若

$$Ee_n e_n^{\mathrm{T}} > 0 \quad (\forall n\in \mathbf{Z}),$$

则 x 称满秩过程. 由于正交投影 $(x_n|H_{n-1})$ 是由极小化过程

$$\inf_{\gamma\in H_{n-1}}\|x_n-\gamma\|^2$$

而得，它常称为 x_n 的线性最小二乘估计，且新息 e_n 常称为(线性)最小二乘预报误差. 新息的概念在随机过程分解、实现、滤波等方面有重要应用.

卡尔曼滤波(Kalman filtering)　离散时间非平稳滤波问题的递推方法. 维纳滤波较成功地解决了平稳过程的滤波问题，然而在20世纪50年代将其推广到非平稳情形的尝试均未成功，其原因是找不到其状态具有固定的有限维数的滤波器实现. 1960年，卡尔曼(Kalman，R. E.)得到了第一个关于离散时间非平稳滤波问题的递推解. 他考虑了下列重要的特殊情形：被估信号是由一个线性有限维状态空间系统产生，且观测值是信号和噪声之和，而噪声和信号正交. 第二年，即1961年卡尔曼和布西(Bucy，R. S.)发表了连续时间系统相应问题的解. 这些结果和方法在工程实际中，特别是在航空航天工程中得到广泛应用. 下面叙述卡尔曼滤波方法.

设 x 和 w 是 n 维向量过程，且 w 满足

$$Ew_iw_j^{\mathrm T}=Q\delta_{ij}\ \ (i,j\in\mathbf Z_+),$$

$$Q>0,\ \delta_{ij}=\begin{cases}0 & (i\neq j),\\ 1 & (i=j).\end{cases}$$

这两过程由 $x_{k+1}=Fx_k+w_k(k\in\mathbf Z_+)$ 相连，设 v 是 p 维过程，且 $Ev_kv_j^{\mathrm T}=R\delta_{kj}$，$Ew_jv_k^{\mathrm T}=0$，$R>0$. 设观测值 y 和 x，v 的关系是 $y_k=Hx_k+v_k\ (k\in\mathbf Z_+)$，其中假设 x_0 和 (w,v) 是正交的. 设 $\hat x_{k|k-1}\triangleq(x_k|H_{k-1}^y)$ 是基于 H_{k-1}^y 对 x_k 的线性最小二乘估计，其中 $H_{k-1}^y=\mathrm{Span}\{y_0,\cdots,y_k\}$，那么 $\hat x_{k|k-1}$ 可由下列方程递推计算：

$$\hat x_{k+1|k}=F\hat x_{k|k-1}+\Delta_k[y_k-H\hat x_{k|k-1}],$$

$$\Delta_k=FV_kH^{\mathrm T}[HV_kH^{\mathrm T}+R]^{-1},$$

$$V_{k+1}=FV_kF^{\mathrm T}-FV_kH^{\mathrm T}[HV_kH^{\mathrm T}+R]^{-1}\cdot HV_kF^{\mathrm T}+Q,$$

$$\hat x_{0|-1}=Ex_0,$$

$$V_0=E(x_0-\hat x_{0|-1})(x_0-\hat x_{0|-1})^{\mathrm T}.$$

当 (w,v) 是正态白噪声时，

$$\hat x_{k|k-1}=E[x_k|\sigma\{y_0,y_1,\cdots,y_k\}]$$

是最小二乘(或最小方差)估计.

在连续时间情形下，相对应的状态方程和观测方程为下列随机微分方程

$$\mathrm dx=Fx\mathrm dt+\mathrm dw,\ \mathrm dy=Hx\mathrm dt+\mathrm dv,$$

其中 (w,v) 是维纳过程，且

$$E\begin{pmatrix}w_t-w_{t_0}\\ v_t-v_{t_0}\end{pmatrix}[w_s^{\mathrm T}-w_{t_0}^{\mathrm T},v_s^{\mathrm T}-v_{t_0}^{\mathrm T}]=\begin{pmatrix}Q&0\\0&R\end{pmatrix},\quad t_0<\min(s,t),$$

$$Ex_{t_0}w_t^{\mathrm T}=Ex_{t_0}v_t^{\mathrm T}=0\ \ (s,t\geqslant t_0).$$

这里假定 x_0 与 (w_t,V_t) 具有联合正态分布，$t\geqslant t_0$，这时对 x_t 的最小方差估计 $\hat x_t$ 可由下式产生

$$\mathrm d\hat x_t=F\hat x_t\mathrm dt+\Delta_t[\mathrm dy_t-H\hat x_t\mathrm dt],$$

$$\Delta_t=V_tH^{\mathrm T}R^{-1},$$

$$\frac{\mathrm dV_t}{\mathrm dt}=FV_t+V_t^{\mathrm T}F^{\mathrm T}-V_tH^{\mathrm T}R^{-1}HV_t+Q,$$

其中 $t\geqslant t_0$，$V_{t_0}\triangleq E(x_{t_0}-\hat x_{t_0})(x_{t_0}-\hat x_{t_0})^{\mathrm T}$.

白噪声(white noise)　一种特殊的平稳随机噪声过程. 它是一种均值为零、谱密度函数为非零的常数，即这种平稳随机过程的功率谱密度 $S(\omega)=$ 常数，也就是均匀分布于整个频率区间 $(-\infty<\omega<\infty)$. 这同自然界的白光相似，白噪声即因此得名. 这在实际物理系统中是不存在的，只是一种数学上的抽象而理想化的随机噪声模型，因为这相当于要求噪声的总功率为

$$\int_{-\infty}^{\infty}S(\omega)\mathrm d\omega=\infty.$$

但是白噪声概念在随机系统的分析和应用中具有非常重要的意义，它类似于确定性系统分析方法中用狄喇克 δ 函数，可以使得随机系统的分析变得简单和方便. 与白噪声过程类似，白噪声序列是白噪声过程的一种离散形式，即如果随机序列 $\{w(k)\}$ 是两两不相关的，相应的自相关函数为

$$R_w(l)=\sigma^2\delta_l\ \ (l=0,\pm1,\pm2,\cdots),$$

其中 $\sigma^2=$ 常数，δ_l 为克罗内克符号，即

$$\delta_l=\begin{cases}1, & 若\ l=0,\\ 0, & 若\ l\neq0,\end{cases}$$

则称随机序列 $\{w(k)\}$ 为白噪声序列. 根据离散傅里叶变换可知白噪声序列 $\{w(k)\}$ 的谱密度为常数 σ^2.

预白化滤波器(pre-whitening filter)　把相关随机序列变换为不相关序列的可逆线性变换. 相关随机序列 $\{y_i\}$ 经过线性变换 W，得到一个"等价"的不相关序列 $\{\varepsilon_i\}$，即设

$$Y^{\mathrm T}=[y_0,y_1,\cdots,y_N],$$

$$\varepsilon^{\mathrm T}=[\varepsilon_0,\varepsilon_1,\cdots,\varepsilon_N],\ \varepsilon=WY.$$

如果 W 是可逆的(即是可逆变换)，W 就是一个预白化滤波器. $\{\varepsilon_i\}$ 称为 $\{y_i\}$ 的新息序列，其中 $\varepsilon_i=y_i-\hat y_{\frac{i}{i-1}}$，$\hat y_{\frac{i}{i-1}}$ 是利用 $\{y_0,y_1,\cdots,y_{i-1}\}$ 对 y_i 的线性最小二乘估计，亦即是 y_i 在 $\{y_0,y_1,\cdots,y_{i-1}\}$ 所张成的空间上的投影. 可见 ε_i 是 $\{y_0,y_1,\cdots,y_{i-1}\}$ 的线性组合 $(i\geqslant0)$，反之 y_i 则是 $\{\varepsilon_0,\varepsilon_1,\cdots,\varepsilon_i\}$ 的线性组合(对于 $i\geqslant0$)，这也就是因果性和因果可逆性. 实质上，相关序列 $\{y_i\}$ 经过 W 变换成不相关序列 $\{\varepsilon_i\}$，即成为两两正交的基. W 的计算方法可用格拉姆-施密特或改进的格拉姆-施密特正交化方法. 当然还可以有许多别的正交化方法(如格拉姆或施密特方法)，这相当

于不同的预白化滤波器的实施方法. 对于连续情形有类似的结果：如果平稳随机过程$\{y(t)\}$的谱密度函数是有理的，根据谱表示定理，则存在一个稳定的线性定常动态系统，其输入是白噪声过程$\{e(s)\}$，其输出为与$\{y(t)\}$具有同样的谱密度函数，即

$$y(t)=\int_{-\infty}^{t}h(t-s)e(s)\mathrm{d}s.$$

当取此线性变换为可逆的(注：存在且惟一)，则可以使$y(t)$经过一个线性滤波器(取作$h(\cdot)$的逆)输出为白噪声的过程$\{e(t)\}$，即

$$e(t)=\int_{-\infty}^{t}g(t-s)y(s)\mathrm{d}s,$$

这里$g(\cdot)$即为预白化滤波器的脉冲响应函数.

最优非线性滤波(optimal non-linear filtering) 非线性随机过程的一种滤波方法. 是在给定观测过程$\{y_t\}$的条件下，做出不可直接观测但与$\{y_t\}$有关联的信号过程$\{x_t\}$的某种误差指标最小的估计值，例如最小方差估计$E\{x_t|y_s,0\leqslant s\leqslant t\}$. 众所周知，对于一切滤波问题，其完整解就是在$y_t\triangleq\sigma\{y_s,0\leqslant s\leqslant t\}$条件下的$x_t$的条件概率密度函数$\pi_t(x)\triangleq p(x_t|y_t)$，这是因为它包括了可利用观测过程中关于$x_t$的一切统计信息. 当观测方程和信号过程方程均是线性的，且噪声是高斯的(包括信号的初始分布也是高斯的)，则$p(x_t|y_t)$是高斯的，由均值和协方差函数完全确定. 因此，计算这些密度比较简单，退化的求解条件均值和条件协方差所满足的方程，这就是卡尔曼滤波方程，它是有限维的.

当观测方程或信号过程方程两者或任一为非线性的，或噪声不是高斯的，则落入最优非线性滤波范畴. 而对于一般非线性滤波情况，不存在描述条件密度的有限参数集，而需要整个地知道这些函数$p(x_t|y_t)$. 如果不对过程$\{x_t\}$假设附加结构条件，人们将得不出什么有用的性质. 如果信号过程$\{x_t\}$是马尔可夫扩散过程，即

$$\mathrm{d}x_t=f(x_t,t)\mathrm{d}t+G(x_t,t)\mathrm{d}\beta_t,\ x_{t_0}=x_0,$$

其中$\{\beta_t\}$为维纳过程(描述布朗运动的随机过程)，其协方差函数为$\theta(t)$；观测过程为

$$\mathrm{d}y_t=h(x_t,t)\mathrm{d}t+\mathrm{d}v_t,$$

其中$\{v_t\}$是具有单位协方差函数的维纳过程，则此时条件概率密度函数的演化满足库希纳方程

$$\mathrm{d}\pi_t(x)=A_1(\pi_t)\mathrm{d}t+B_1(\pi_t)\mathrm{d}y_t,$$

其中A_1,B_1分别是积分和微分算子符. 库希纳方程是非线性随机偏微分方程，一般是无限维的. 如果采用非归一化条件概率密度函数$\rho_t(x)$，使

$$\pi_t(x)=\frac{\rho_t(x)}{\int_{-\infty}^{\infty}\rho_t(x)\mathrm{d}x}$$

(“非归一化”概率密度函数的意思指

$$\int_{x}^{\infty}\rho_t(x)\mathrm{d}x$$

不一定等于1)，则此时可得与库希纳方程等价的非归一化变型，即具有更简单形式的扎凯方程

$$\mathrm{d}\rho_t(x)=A_2\rho_t(x)+h\rho_t(x)\mathrm{d}y_t.$$

它是一个线性(对ρ_t来说)随机偏微分方程. 在一般情况下，扎凯方程仍然是不容易求解的，因此非线性滤波器除去少数特殊情况(例如，卡尔曼滤波器、本耐思滤波器以及能覆盖两者的但姆滤波器等)存在精确的有限维解以外，一般都是无限维的，只能近似求解.

扎凯方程(Zakai equation) 一种随机偏微分方程. 指扎凯(Zakai, M.)在求解下述非线性滤波问题时，得到的双线性随机偏微分方程. 设

$$\mathrm{d}x_t=f(x_t,t)\mathrm{d}t+g(x_t,t)\mathrm{d}w_t,$$
$$\mathrm{d}y_t=h(x_t,t)+\mathrm{d}v_t\quad(0\leqslant t\leqslant T),$$

其中$\{w_t\}$和$\{v_t\}$的均值为零，协方差函数分别为$q(t)$和单位值的维纳过程，且它们之间与初值x_0(高斯分布)相互独立. x_t的最小方差(最优)估计是

$$\hat{x}_t=E\{x_t/y_s;0\leqslant s\leqslant t\}.$$

条件概率密度函数

$$p(x_t/y_t)=\pi_t(x);\ y_t\triangleq\sigma(y_s;0\leqslant s\leqslant t\}$$

满足非线性随机偏微分方程而随时间变化. 如果定义非归一化概率密度函数$\rho_t(x)$满足

$$\pi_t(x)=\frac{\rho_t(x)}{\int_{-\infty}^{\infty}\rho_t(x)\mathrm{d}x}$$

(“非归一化”概率密度函数的意思指$\int_{x}^{\infty}\rho_t(x)\mathrm{d}x$不一定等于1)，则$\rho_t(x)$可通过扎凯方程

$$\mathrm{d}\rho_t(x)=A_2\rho_t(x)+h\rho_t(x)\mathrm{d}y_t$$

求得，其中A_2是柯尔莫哥洛夫算子，

$$A_2=\frac{\partial}{\partial x}\left(-f+\frac{1}{2}gqg\frac{\partial}{\partial x}\right).$$

这是一个线性(对于$\rho_t(x)$，而对输入驱动$\mathrm{d}y_t$是双线性)的随机偏微分方程.

推广的卡尔曼滤波(extended Kalman filter) 最优线性卡尔曼滤波在非线性动态系统中的推广. 其基本思想是把该非线性方程围绕当前状态估值线性化(取泰勒级数展开式的一阶近似式)，然后再对此近似的线性方程使用卡尔曼滤波公式. 现以离散时间系统为例，具体说明如下：

$$x_k=f_{k-1}(x_{k-1})+\Gamma_{k-1}(x_{k-1})w_{k-1},\qquad(1)$$
$$y_k=h_k(x_k)+v_k,\qquad(2)$$

其中$w_k\in N(0,Q_k)$，$v_k\in N(0,R_k)$，$x_0\in N(\mu_0,P_0)$. 又$\{v_k\}$与$\{w_k\}$及x_0都相互独立. 上述(1)，(2)方程中的f围绕$\hat{x}_{k-1}$展成泰勒级数，并取一阶近似，又用$\Gamma_{k-1}(\hat{x}_{k-1})$代替$\Gamma_{k-1}(x_{k-1})$，则得到(1)式的线性化模型

$$x_k \approx \frac{\partial f_{k-1}}{\partial \hat{x}_{k-1}} x_{k-1} + \left[f_{k-1}(\hat{x}_{k-1}) - \frac{\partial f_{k-1}}{\partial \hat{x}_{k-1}} \hat{x}_{k-1} \right] + \Gamma_{k-1}(\hat{x}_{k-1}) w_{k-1}. \quad (1)'$$

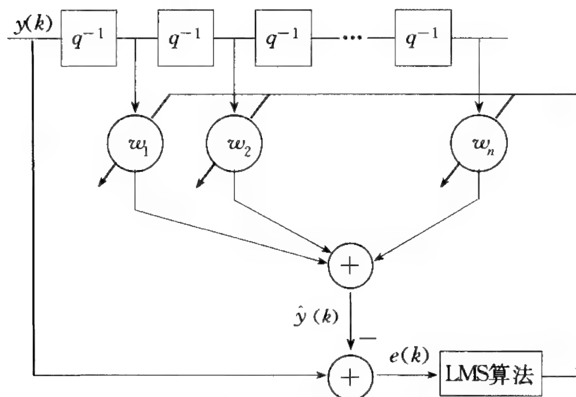
再把观测方程(2)中的 h 围绕 $\hat{x}_{k|k-1}=f_{k-1}(\hat{x}_{k-1})$ 展开成泰勒级数,而取其线性项,可得

$$z_k \approx \frac{\partial h_{k-1}}{\partial \hat{x}_{k|k-1}} x_k + \left[h_{k-1}(\hat{x}_{k|k-1}) - \frac{\partial f_{k-1}}{\partial \hat{x}_{k|k-1}} \hat{x}_{k|k-1} \right] + v_k. \quad (2)'$$

然后再对线性化后的(1)′,(2)′应用卡尔曼滤波公式就得到推广的卡尔曼滤波.

必须注意,这时线性化是在每得到一个新的估值之后才能进行,才能得到各系统系数矩阵,它必须实施在线计算.这与常规的卡尔曼滤波是不同的,因为估计误差协方差的离散黎卡提矩阵方程与状态估计方程相互不独立,因此,协方差矩阵和增益矩阵不可能在知道状态估计值之前事先算好,而也都必须实施在线计算.

最小均方适应滤波器(least mean squares adaptive filter) 能求解自适应滤波器系数的一种方法.由维佐夫(Widrow,B.)和霍夫(Hoff,M. E. Jr.)提出的这一算法利用最速下降原理,由均方误差的梯度估计从现时刻滤波器系数向量迭代计算下一时刻的系数向量,实质上是最小均方误差算法(简称 LMS 算法).由于 LMS 算法简单易行,它是获得最广泛应用的一种自适应算法,用下图说明之.图中



q^{-1}是延迟一拍的算子,LMS 算法为

$$W(k+1) = W(k) - \mu \nabla_W E\{e^2(k)\},$$

其中 $W(k)=[w_1, w_2, \cdots, w_n]^{\mathrm{T}}$,T 表示转置. $\nabla_W E\{e^2(k)\}$表示均方误差对 W 系数的偏导,实际计算时用$\nabla_W\{e^2(k)\}$代之.μ 为一常数,表示步长,它的取值决定算法的收敛性,一般要求

$$0 < \mu < \frac{1}{\lambda_{\max}},$$

$\lambda_{\max}$为输入信号序列$\{y(k)\}$自相关矩阵最大特征值.

最小均方适应滤波器成功地应用于自动均衡、回波消除、噪声消除、天线阵波束形成、语音信号处理和谱估计等方面.

随机控制(stochastic control) 一类控制方法.受到随机因素影响的系统即随机系统.对随机系统最常用的数学描述是随机微分方程或随机差分方程,这取决于连续时间还是离散时间.下面来看离散时间的随机系统

$$x_{k+1} = f_k(x_k, u_k, \xi_{k+1}), \quad (1)$$

$$y_k = g_k(x_k, u_{k-1}, \eta_k), \quad (2)$$

x_{k+1}是系统在 $k+1$ 时刻的状态,它是 k 时刻的状态 x_k、控制 u_k 及随机噪声 ξ_{k+1}的函数.方程(1)称为状态方程. y_k 是 k 时刻的量测,当 $y_k=x_k$ 时,系统称为完全状态信息的.但通常 x_k 不能直接量测到,而量到的只是 x_k,u_{k-1}及随机噪声 η_k 的函数 $g_k(x_k, u_{k-1}, \eta_k)$.当 f_k 和 g_k 都是线性函数时,则系统称为线性随机系统.它的典型形式是

$$x_{k+1} = A_k x_k + B_k u_k + D_k w_{k+1}, \quad (3)$$

$$y_k = C_k x_k + G_k u_{k-1} + F_k w_k, \quad (4)$$

其中 $A_k, B_k, C_k, D_k, G_k, F_k$ 是确定性阵,$w_k^{\mathrm{T}}=[\xi_k^{\mathrm{T}}, \eta_k^{\mathrm{T}}]$.当$\{w_k\}$是相互独立的正态序列时,系统(3),(4)称为高斯系统.这时如果(3)中的 A_k, B_k, D_k 依赖$(y_0, y_1, \cdots, y_k)$,而(4)中的 C_k, G_k, F_k 依赖于$(y_0, y_1, \cdots, y_k)$,那么系统(3),(4)称为条件高斯系统.

u_k 是 k 时刻的控制量,而在 k 时刻可用的信息只有$(y_0, y_1, \cdots, y_k)$,所以 u_k 是$(y_0, y_1, \cdots, y_k)$的函数,也就是 u_k 是反馈控制.控制的目的是要使性能指标 $J(u)$最小.

$$J(u) = E\left[\sum_{i=1}^{N-1} L_i(x_i, u_i) + L_N(x_N)\right], \quad (5)$$

N 是终端时刻,E 表示数学期望.使 $J(u)$达最小的反馈控制$\{u_k\}$就是系统(1),(2),(5)或(3),(4),(5)的最优随机控制.但能得到显式表达的最优随机控制的系统只是个别情形,其中一个重要的系统就是 $J(u)$是二次指标时的线性高斯系统(参见“LQG 问题”).

高斯系统(Gauss systems) 见“随机控制”.

随机优化准则(stochastic optimization criterion) 求随机系统中最优控制的一个准则.和确定性系统相似,对随机系统有随机极大值原理和动态规划.动态规划的依据称优化准则:无论初值状态和以前的控制取何值,余下的控制相对于由过去控制所产生的状态,必须是最优的.如果用

$$J_k = E\left[\sum_{i=k}^{N-1} L_i(x_i, u_i) + L_N(x_N)\right]$$

表示在 k 时刻到终点的价格,那么根据优化准则,从终点时刻倒着往回走,对每一个时刻 k 都要取 u_k 使 J_k 达到最小,这样就得到最优随机控制.确切地说,要倒着递推地定义“值函数”$V_k(x)$($k=N, N-1$,

…)，首先 $V_N(x)=L_N(x)$，在 $N-1$ 时刻求 u_{N-1}，使 $J_{N-1}=L_{N-1}(x_{N-1},u_{N-1})+V_N$ 达极小，极小值记为 $V_{N-1}(x_{N-1})$. 对任意时刻 $k(1\leqslant k\leqslant N-1)$，要求 u_k 使

$$V_k(x_k) = \inf_{u_k}[L_k(x_k,u_k)+E_\xi V_{k+1}(f_k(x_k,u_k,\xi_{k+1}))],$$

这里 E_ξ 表示只对 ξ_{k+1} 求期望. 如果能求出这样的序列 $\{u_k\}$，那么这就是根据随机优化准则求出的最优随机控制.

分离原则(separate principle)　状态估计和求最优控制可分开进行的原则. 对“随机控制”条目中的(1)，(2)或(3)，(4)讨论性能指标(5)时，要注意性能指标依赖于状态变量 $\{x_k\}$，但 x_k 不能直接量测到，所以要从量测信息 $(y_0,y_1,\cdots,y_k)$ 求对 x_k 的估计，即滤波. 所以最优随机控制蕴含着两方面的问题：一是对状态滤波，二是求控制使性能指标达到极小. 如果最优随机控制问题可以分开来解决：先求出对状态 x_k 的滤波值 $\hat{x}_k$，然后把相应的确定性系统(噪声恒为零)的最优控制中的 x_k 用 $\hat{x}_k$ 来取代，就得到最优随机控制，那么就称成立分离原则，或称必然等价原则. 对“随机控制”条目中的(1)，(2)，一般说 $\hat{x}_k$ 不能用一组封闭的方程来决定(条件正态系统是一个重要的例外)，成立分离原则就更不容易了，但对实际应用中十分重要的 LQG 问题可用成立分离原则.

线性二次高斯问题(linear quadratic Gaussian problem)　简称 LQG 问题. 是讨论高斯系统(参见“随机控制”中系统(3)，(4))在极小化二次型性能指标的问题

$$J(u)=E\left\{x_N^{\mathrm{T}}Q_0x_N+\sum_{i=k}^{N-1}(x_i^{\mathrm{T}}Q_1x_i+u_i^{\mathrm{T}}Q_2u_i)\right\},$$

这里 $Q_0\geqslant 0,Q_1\geqslant 0,Q_2>0$ 分别是对称半正定和对称正定阵. 当“随机控制”条目中的(3)中无噪声(即 $w_{k+1}=0$)，且状态完全量测时，相应的性能指标也不再取期望，这时称之为线性二次问题. 这时最优控制是状态量的线性反馈 $u_k=L_kx_k$，L_k 由“随机控制”条目中的(3)中的系数阵及 Q_0,Q_1,Q_2 所决定，它可以递推地表达出来. 对高斯系统，滤波值 $\hat{x}_k$ 可以递推地求得，这就是卡尔曼滤波. 把滤波值 $\hat{x}_k$ 取代相应确定性系统最优控制中的状态 x_k，得到随机控制 $u_k=L_k\hat{x}_k$. 这就是使二次型指标达极小的最优随机控制. 所以对 LQG 问题成立分离原则.

随机李亚普诺夫函数(stochastic Liapunov function)　研究随机动态系统渐近性质的工具. 它是帮助人们在不求解随机微分方程的情况下，定性地研究该随机系统的渐近性态. 例如，随机稳定性、各态历经性、循环、瞬态、平稳性等，这些概念都是用于研究随机系统的渐近特性、涉及状态过程样本函数是否以概率 1 收敛到状态空间的某一子空间. 此时把李亚普诺夫函数的概念推广于随机系统作为定性分析手段，是有力的、合适的工具. 随机李亚普诺夫函数定义如下：设 $U\subset\mathrm{R}^n$ 是一个有界闭域，标量函数 $v(t,x)\in C_0^{1,2}(\{t\geqslant t_0\}\times U)$，即在定义域 $\{t\geqslant t_0\}\times U$ 上，$v(t,x)$ 对 t 有一阶连续偏导数 $\partial v/\partial t$，对于 x 有二阶连续偏导数

$$\frac{\partial v}{\partial x},\quad \frac{\partial^2 v}{\partial x^2}.$$

当 $x=0$ 时，$v(t,0)\equiv 0$，并且，对于一切不等于零的 $x\in U$ 时

$$v(t,x)>0, \tag{1}$$

$$yv(t,x)\leqslant 0, \tag{2}$$

其中，y 为过程 $\{x(t)\}$ 的微分发生器算子. 可定义满足上述条件(1)，(2)的函数 $v(t,x)$ 为随机李亚普诺夫函数.

线性最小方差控制(linear minimum variance control)　线性随机系统的一种优化控制方法. 它是根据给定对象的数学模型，设计出一个线性控制律，使输出的稳态均方差为最小. 设考虑对象过程

$$A(q^{-1})y(t)=q^{-d}B(q^{-1})u(t)+C(q^{-1})\xi(t),$$

其中

$$A(q^{-1})=1+a_1q^{-1}+\cdots+a_{n_a}q^{-n_a},$$

$$B(q^{-1})=b_0+b_1q^{-1}+\cdots+b_{n_b}q^{-n_b},$$

$$C(q^{-1})=1+c_1q^{-1}+\cdots+c_{n_c}q^{-n_c},$$

$$q^{-1}y(t)=y(t-1),$$

$\xi(t)$ 为白噪声序列，即

$$E\{\xi(t)\}=0,\ E\{\xi(i)\xi(j)\}=\begin{cases}\sigma^2 & (i=j),\\ 0 & (i\neq j).\end{cases}$$

过程的 $B(q^{-1})$ 和 $C(q^{-1})$ 是零点在单位圆外的多项式，此时使过程实际输出 $y(t+d)$ 与希望输出 $y_m(t+d)$ 之间的误差的均方差

$$J=E\{[y(t+d)-y_m(t+d)]^2\}$$

达到最小的控制律为

$$B(q^{-1})F(q^{-1})u(t)=-G(q^{-1})y(t),$$

其中

$$F(q^{-1})=1+f_1q^{-1}+\cdots+f_{d-1}q^{d-1},$$

$$G(q^{-1})=g_0+g_1q^{-1}+\cdots+g_{n-1}q^{n-1},$$

满足下列丢番图方程

$$C(q^{-1})=A(q^{-1})F(q^{-1})+(q^{-d})G(q^{-1}).$$

输出量的误差 $\tilde{y}(t)$ 是 $e(t)$ 的 $d-1$ 阶滑动平均，即 $\tilde{y}(t)=F(q^{-1})e(t)$. 输出量的误差最小方差等于

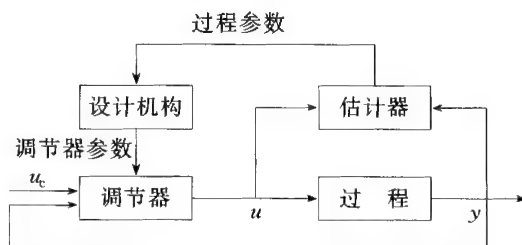
$$E\{\tilde{y}^2(t)\}=(1+f_1^2+\cdots+f_{d-1}^2)\sigma^2.$$

最小方差控制只能用于逆稳定的受控对象，这是其基本缺点. 关于多输入、多输出系统的最小方差控制问题，也有相应的结果.

适应控制(adaptive control)　一类控制方法.能在系统和环境改变或信息不全的情形下,改变自身的特性来保持系统的良好品质的控制,又称"自适应控制".具体地说,适应控制通过对系统和环境不断地观测,对已有的控制效果进行分析,自动地调整控制器结构或参数,达到不断改善系统品质的目的.初期用模型参考适应控制来设计飞机的自动驾驶仪,这种控制是把系统的实际输出与理想模型的输出进行比较,自动调整控制器的参数使得系统的实际输出与理想模型的输出保持一致.后来自校正适应控制又兴起了,自校正控制是当系统特性和环境未知或变化时,用量测数据实时地识别这些变化并自动地修改参数,使系统达到期望的指标.闭环适应控制系统一般是由高度非线性的动态方程组所描述,它的稳定性和鲁棒性是重要的研究课题.

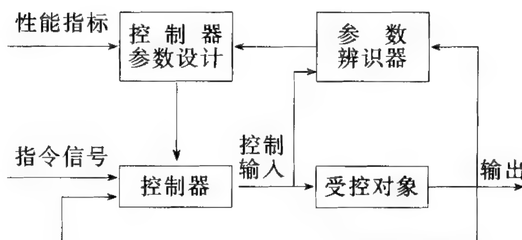
预测控制(predictive control)　一类特定的控制方法.它是利用预测模型和系统的历史数据、未来输入来预测系统的未来输出的控制.通过某一性能指标在滚动的有限时间区间内进行优化得到反馈校正控制.状态方程、传递函数及稳定系统的阶跃响应、脉冲响应函数等都可以作为预测模型.滚动优化与传统的全局优化不同,滚动优化在每一时刻优化性能指标只涉及从该时刻起未来有限的时间,而到下一时刻,这一优化时间同时向前推移,不断地进行在线优化.在每一时刻得到一组未来的控制动作,而只实现本时刻的控制动作,到下一时刻重新预测优化出一组新的控制,也是只实现一个新的控制动作,每步都是反馈校正.预测控制有了预见性,滚动优化和反馈校正能够更好地适应实际系统,有更强的鲁棒性.预测控制主要有动态矩阵控制、模型算法控制和广义预测控制.

自校正调节器(self-tuning regulator)　是实现适应控制的一种调节器.在一个自适应系统中,调节器的参数需要不断进行调整以跟踪过程的变化.不妨设过程参数是恒定且未知的,当过程已知时,设计过程规定了一组希望的控制器参数;而当过程未知时,自适应控制器的参数应当收敛到这些希望的参数值.具有这种性质的调节器称为自校正调节器,这是因为它能自动校正自己的参数以得到希望的闭环系统性能.自校正调节器的设计思想是把未知参数的估计和控制器的设计分开进行,其框图如下.未知参数用递推估计方法在线估计.各种不同的估计方法都可用来估计未知参数,如随机逼近法、最小二乘法、辅助变量法和极大似然法等.下图中的设计机构方框表示在线求解参数已知系统的控制设计问题,常用的方法有最小方差、线性二次型、极点配置和模型跟踪等方法.不同的估计方法和设计方法的组合可导出性质不同的调节器,自校正调节器原先是从采样数据系统导出,现在已研究出连续系统的自校正算法和混合校正算法.



必然等价原则(certainly equivalent principle)　线性系统辨识中参数估计常用的一个原则.用递推估计方法在线估计未知参数,把估计出的参数看成是真参数,用到控制律中去,而不考虑估计的不定性,这种原理称为必然等价原则或确定性等价原理.

显式适应控制(explicit adaptive control)　一种基于先辨识系统参数再构造控制器的适应控制方法.一个自适应控制系统必须不断地提供受控对象未知参数,也就是必须不断进行辨识,然后将系统的当前性能与期望的或最优的性能进行比较,做出使系统趋向期望的性能的决策.根据该决策对控制器进行设计和不断修正,以驱使系统达到最优状态.自校正控制系统如图所示.



这种不断根据受控对象的输入、输出来辨识对象的数学模型,然后再针对性能指标要求及当时辨识的参数模型设计和计算出控制器的参数,称为显式适应控制算法.由于这种算法是先辨识受控系统的参数,再计算控制器参数,故又称为间接法.

间接适应控制(indirect adaptive control)　即"显式适应控制".

隐式适应控制(implicit adaptive control)　一种使用输入、输出观测数据直接估计控制器参数的适应控制方法.一个自适应控制系统必须根据受控对象及周围环境的变化不断修正调整控制动作,以使系统趋向期望的或某一指标意义下的最优状态.在显式(间接)适应控制中是先用被控对象的输入、输出数据辨识该对象数学模型的参数,然后再按照性能指标要求计算出控制器的参数(参见"显式适应控制").另一种使用输入、输出观测数据的方式是直接估计控制器的未知参数.这时需要将过程重新参

数化，即建立一个与控制器参数直接关联的估计模型. 由于这时没有受控对象的数学模型参数的辨识过程，而直接估计控制器参数，故称隐式适应控制，或称直接适应控制. 隐式(直接)算法无需进行控制器参数的计算，所以它的计算量比显式算法要小，不过需要为它建立一个合适的控制器参数的估计模型. 在自校正调整器中大多数采用这种方法，这是因为隐式算法不仅比显式算法简单，而且鲁棒性也更强些.

直接适应控制(direct adaptive control)　即"隐式适应控制".

适应零极配置(adaptive pole-zero placement)　自适应控制系统的一种综合设计方法. 当多输入、多输出系统的传递函数阵未知时，设计控制器，使闭环系统的参考输入到误差的传递函数具有全部希望的极点和部分希望的零点配置. 众所周知，不仅系统的稳定性取决于极点的分布，而且系统控制品质(例如上升时间、超调量、振荡次数、建立时间等)在很大程度上也是与极点的位置密切相关. 但系统的跟踪特性特别是跟踪高阶指令信号，则与零点的配置有很大关系，因此对于要求系统动态特性又要求跟踪特性好的系统就有必要寻找同时适应零极配置的设计方法. 上面提到部分希望的零点配置是因为任意配置闭环零点，就必须对消过程的全部零点，然而采样一个连续对象而得到的离散模型往往是逆不稳定的，所以无法满足任意配置闭环零点的条件.

自校正控制器(self-tuning controller)　又称自校正跟踪器. 基于自校正调节器发展起来的一种控制器. 在自校正调节器中，认为伺服输入为零，所讨论的优化指标函数是输出的最小方差，不包括伺服输入项，也没有考虑对控制项的约束(这样的控制作用可能超出允许值很多). 为了克服这些缺点，发展了自校正控制器. 这时输出跟踪伺服输入，这种控制器算法仍采用二次型指标函数，但在指标函数中引入伺服输入项和控制作用加权项，即形式为

$$J = E\{[y(t) - y^*(t)]^2 + [\lambda u(t)]^2\},$$

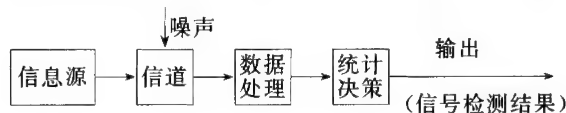
其中 $y(t)$ 是输出项，$y^*(t)$ 是伺服输入项，$u(t)$ 是控制项，λ 是加权系数. 因此，它除了多一个伺服(跟踪)功能外，而且限制了控制信号的不适当增长. 同时，适当选择控制加权因子大小，可以使之适用于非逆稳定(非最小相位)系统.

随机实现(stochastic realization)　把宽平稳过程表达成以正交平稳过程为输入的线性系统的输出. 设 y 是 p 维宽平稳随机过程，方差函数为 $\{R_i|i\in \mathbb{Z}\}$，$z=\{z_i: z_i\in \mathbb{R}^{p\times q}, i\in \mathbb{Z}\}$ 是线性时不变系统的脉冲响应序列，并设 w 是 q 维宽平稳正交序列. 如果系统 z 在输出 w 下是宽平稳过程，且具有方差函数 $\{R_i|i\in \mathbb{Z}\}$，那么 $\{z,w\}$ 称为 y 的弱(输入-输出)随机实现.

设 y 和 w 定义于同一概率空间 $(\Omega, \mathscr{F}, P)$，如果 y 和 w 的样本轨道满足 $y=zw$ 几乎处处成立，那么 $\{z,w\}$ 称为 y 的强(输入-输出)随机实现. 进一步，如果 w 是 y 的新息过程，那么 $\{z,w\}$ 称为 y 的(输入-输出)新息实现. 类似地，可以定义强(弱)状态空间随机实现. 沃尔解分解定理说明，对于宽平稳纯线性非确定性随机过程都具有新息实现. 此外还可以直接从 y 的谱密度矩阵的任何谱因子得到宽平稳纯线性非确定性随机过程的弱随机实现.

信号检测(signal detection)　从带噪声的数据中判断某种信号存在的问题. 带有信息的各种形式的可观测、可传输、可存储的信号均以数据形式表示，这种数据在传输、存储和变换过程中不可避免地会受到噪声或其他无用数据的干扰. 在接收到的有噪声的数据中判断是否有某种信号存在，就是信号检测问题. 例如，要在雷达的输出数据中，判断在警戒空域中是否存在敌方飞机的信息，就是典型的信号检测例子.

20 世纪 50 年代起，人们开始将统计推断理论中的假设检验、参数估计、统计判决及序列分析等统计数学工具用于信号检测问题，建立起信号检测的统计理论. 信号检测的过程如下图所示. 如果信息源发出的原始信息只有两种可能：H_1(表示有)或 H_0



(表示没有)，统计判决过程只是在 H_0 和 H_1 中按某种判决准则(例如，最大似然比等)选择一种，这称为二元检测，属较简单的情况. 如果原始数据中有多种可能，$H_1, H_2, \cdots, H_n$，统计判决过程则需要在几种假设中选择出一种，称为多元检测.

马氏决策(Markov decision)　一类可连续进行观察的随机动态系统的最优化决策. 根据在各时刻观察到的状态，从允许的决策(控制、措施等)集合中选用一个决策，在这里假定状态和决策两者都不依赖于系统的过去的历史. 在各个时刻选取决策的目的，是使系统运动的全过程达到某种最优运行效果的决策规则，称为马氏决策. 即每个时刻 t 选取决策的规则，即与系统在 t 以前的历史无关，又是与当前系统状态 i 所对应的决策 $f_t(i)$，$i\in S$，S 是系统全体状态所成的集合. 若各个时刻的决策规则都是相同的马氏决策，则称为平稳决策，即$(f,f,f,\cdots)$(记为 f^∞). 由于平稳决策的简单性，因此它是实际应用中特别重要的一类决策. 马氏决策过程是动态规划与马尔可夫过程结合的产物，由贝尔曼(Bellman, R.) 1957 年正式提出并借助于最优性原理求解最优决

策的方法.后来,有人发现寻求最优决策问题可以化为求解相应的线性规划问题.目前,它已在设备的更换与维修、库存论、排队论、控制工程、可靠性理论、搜索论、水库调度、林渔业管理、通讯网络等的最优化问题中都有应用,并正向工程、生物、经济等领域渗透.

研究者注意的问题主要有:各类模型及其更一般化的研究;状态部分可观察、适应性等模型的理论探讨;特殊模型更有效的解法;如何用易于处理的模型去逼近复杂的模型等问题.

分布参数系统理论

分布参数系统理论(theory of distributed parameter system) 研究无穷维系统的品质、建模、控制等理论的总称.分布参数系统通常是指由偏微分方程、积分方程、泛函微分方程等描述的动力学系统,与有穷维系统相比,无穷维系统有其自身特有的问题,处理问题的数学工具主要是泛函分析、偏微分方程理论等.

自从20世纪60年代初现代控制理论诞生,分布参数系统理论的研究随之也就开始了.最早从事这一领域研究的学者有布特柯夫斯基(Butkovskiy, A. G.)和勒奈夫(Lernev, A. Y.)(1960)、王耿介(1964)、莱昂斯(Lions, J.-L.)(1966)和布洛根(Brogan, W. L.)(1967)等.

目前已有大量文献和报告涉及分布参数系统的建模、稳定性、仿真、辨识、估计和控制等理论问题,以及包括化工系统、机械系统、环境系统、等离子体控制、航空航天等在内的众多重要的应用.

预计分布参数系统研究的重点将逐步从线性模型转向更加复杂的非线性模型,如在非线性弹性力学和流体力学中出现的重要的非线性系统.但由于无穷维系统的复杂性,对于这种系统的控制问题的深入研究似乎需要在概念上有一些重大的突破.当然,在线性分布参数系统的控制方面仍还有许多问题需要解决,例如,需要开展无穷维线性系统的鲁棒控制理论的研究.如果能将有穷维系统中已经取得的一些结果推广到分布参数系统,则很有可能导致无穷维系统的十分有用的控制方法的出现.实际上,有穷维控制理论,如能控性、能观测性、最优控制、稳定性、镇定、适应控制、非线性控制、鲁棒性等,在分布参数系统控制中仍然是十分重要的,有的甚至是相当活跃的课题.而且正是由于系统无穷维特性,才出现新的难度大的数学问题,吸引着众多的数学工作者的研究兴趣.

就偏微分方程描述的系统的控制来说,计算上面临着挑战:

1. 如何化成有穷维问题.
2. 由空间和时间离散化导致庞大的有穷维系统.
3. 方程的定义域随时间变化而出现新的困难.

目前刺激分布参数系统研究的几个主要的应用领域有:

1. 大型柔性空间结构.
2. 柔性机器人.
3. 燃烧控制.
4. 流体动力控制.
5. 结构和材料的优化设计.
6. 水资源管理.
7. 石油勘探和开采等.

分布参数系统(distributed parameter system) 一个动力学系统.是由偏微分方程、泛函方程、积分-微分方程、积分方程以及抽象空间中微分方程描述的动力学系统.这些系统的共同特征是具有无穷维的状态空间.与有穷维系统不同,分布参数系统具有许多特有的问题,如奇异摄动效应、解的正则性,以及选择适合数值计算的有限维逼近等.现代科学技术的许多领域,诸如与航天有关的大型空间结构、柔性机器人、流体流动、轧钢、人口、社会经济、环境、以及生物医学中的许多问题都归结到分布参数系统的研究.

分布参数控制系统(distributed parameter control system) 含控制作用的分布参数系统.泛指用分布参数系统描述的受控对象的控制,控制的目的是改变系统的性能品质,以适合人们的需要.例如,使不稳定的系统达到稳定,采用的方法可以是反馈镇定,或最优控制(时间最省、能量最省等).控制作用可以是点控制、分布控制、边界控制等.

无穷维控制系统(infinite dimensional control system) 即"分布参数控制系统".但通常多指抽象空间中的控制系统.

无穷维线性控制系统(infinite dimensional linear control system) 无穷维巴拿赫空间或希尔伯特空间X(状态空间)中的线性控制系统.它的形式为

$$\dot{x} = Ax(t) + Bu(t),$$

这里A是X中某个强连续线性算子半群的无穷小生成算子,控制量$u(t)$属于另一个巴拿赫空间或希尔伯特空间U(控制空间),B是U到X的线性算子.许多由偏微分方程描述的控制系统,如抛物型、双曲型线性控制系统都可以化成这种抽象形式来处理.

椭圆型控制系统(controlled system governed by elliptic partial differential equation) 一类由椭圆型偏微分方程描述的分布参数控制系统.设Ω是

$\mathbb{R}^n$ 中的区域，其边界 Γ 是光滑的，Γ 的外法线 $\boldsymbol{n}$ 的第 i 个方向余弦为 $\cos(\boldsymbol{n},\boldsymbol{x}_i)$；$a_0,a_{ij}\in L^\infty(\Omega)$满足条件

$$\sum_{i,j=1}^{n}a_{ij}(x)\xi_i\xi_j \geqslant \alpha\sum_{i=1}^{n}\xi_i^2 \quad (\forall \xi_i \in \mathbb{R}, x \in \Omega),$$

其中 $\alpha>0$ 是常数. 由椭圆型偏微分方程

$$Ay(x) \triangleq -\sum_{i,j=1}^{n}\frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial y(x)}{\partial x_j}\right)+a_0(x)y(x) = f(x)+b(x)u(x) \ (\forall x\in\Omega)$$

和边值条件 $y(x)=0, \forall x\in\Gamma$，或

$$\frac{\partial y(x)}{\partial \nu_A}\triangleq \sum_{i,j=1}^{n}a_{ij}(x)\frac{\partial y(x)}{\partial x_j}\cos(\boldsymbol{n},\boldsymbol{x}_i) = 0, \forall x \in \Gamma$$

描述的受控系统，称为椭圆型控制系统. 这时，控制作用 $u(\cdot)$称为分布控制. 如果 $b(x)\equiv 0$，而边值条件换为

$$\frac{\partial y(x)}{\partial \nu_A} = u(x) \quad (\forall x \in \Gamma)$$

或

$$y(x) = u(x) \quad (\forall x \in \Gamma),$$

控制作用 $u(\cdot)$称为边值控制. 椭圆型控制系统的抽象形式可写为

$$Ay = f + Bu.$$

对于分布控制，$(Bu)(x)=b(x)u(x)$；对于边值控制，B 是映 Γ 上的函数到 Ω 上函数的线性算子，可能是无界算子.

抛物型控制系统(controlled system governed by parabolic partial differential equation) 一类由抛物型偏微分方程描述的分布参数控制系统. 设 Ω 是 $\mathbb{R}^n$ 中的区域，其边界 Γ 是光滑的，$\cos(\boldsymbol{n},\boldsymbol{x}_i)$是 Γ 的外法线 $\boldsymbol{n}$ 的第 i 个方向余弦；$a_0(\cdot), a_{ij}(\cdot)\in L^\infty(\Omega)$满足条件

$$\sum_{i,j=1}^{n}a_{ij}(x)\xi_i\xi_j \geqslant \alpha\sum_{i=1}^{n}\xi_i^2 \quad (\forall \xi_i \in \mathbb{R}, x \in \Omega),$$

其中 $\alpha>0$ 是常数. 由抛物型偏微分方程

$$\frac{\partial y(t,x)}{\partial t}-\sum_{i,j=1}^{n}\frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial y(t,x)}{\partial x_j}\right)+a_0(x)y(t,x) = f(t,x)+b(x)u(t,x), \quad \forall(t,x)\in(0,T)\times\Omega,$$

初值条件 $y(0,x)=y_0(x), \forall x\in\Omega$ 和边值条件

$$\frac{\partial y(t,x)}{\partial \nu_A}\triangleq\sum_{i,j=1}^{n}a_{ij}(x)\frac{\partial y(t,x)}{\partial x_j}\cos(\boldsymbol{n},\boldsymbol{x}_i)=0, \quad \forall(t,x)\in(0,T)\times\Gamma$$

描述的受控系统，称为抛物型控制系统. 它的抽象形式为

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} + Ay(t) = f(t) + Bu(t),$$

这里 $-A$ 是 C_0 算子半群的母元，B 是线性算子. 对于分布控制作用，$(Bu(t))(x)=b(x)u(t,x)$. 对于边值控制作用，B 可能是无界的.

双曲型控制系统(controlled system governed by hyperbolic partial differential equation) 一类由双曲型偏微分方程描述的分布参数控制系统. 设 Ω 是 $\mathbb{R}^n$ 中的区域，其边界 Γ 是光滑的，$\cos(\boldsymbol{n},\boldsymbol{x}_i)$是 Γ 的外法线 $\boldsymbol{n}$ 的第 i 个方向余弦；$a_0(\cdot), a_{ij}(\cdot)\in L^\infty(\Omega)$满足条件

$$\sum_{i,j=1}^{n}a_{ij}(x)\xi_i\xi_j \geqslant \alpha\sum_{i=1}^{n}\xi_i^2 \quad (\forall \xi_i \in \mathbb{R}, x \in \Omega),$$

其中 $\alpha>0$ 是常数. 由双曲型偏微分方程

$$\frac{\partial^2 y(t,x)}{\partial t^2}-\sum_{i,j=1}^{n}\frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial y(t,x)}{\partial x_j}\right)+a_0(x)y(t,x) = f(t,x) + b(x)u(t,x), \quad \forall(t,x)\in(0,T)\times\Omega,$$

初值条件

$$y(0,x) = y_0(x), \quad \left.\frac{\partial y(t,x)}{\partial t}\right|_{t=0} = y_1(x), \forall x \in \Omega,$$

和边值条件

$$\frac{\partial y(t,x)}{\partial \nu_A} \triangleq \sum_{i,j=1}^{n}a_{ij}(x)\frac{\partial y(t,x)}{\partial x_j}\cos(\boldsymbol{n},\boldsymbol{x}_i) = 0 \quad (\forall (t,x) \in (0,T) \times \Gamma)$$

描述的受控系统，称为双曲型控制系统. 它的抽象形式为

$$\frac{\mathrm{d}^2 y(t)}{\mathrm{d}t^2} + Ay(t) = f(t) + Bu(t),$$

这里 A 是解析半群的母元，B 是线性算子. 对于分布控制作用，$(Bu(t))(x)=b(x)u(t,x)$，B 是有界的. 对于边值控制作用，B 可能是无界的.

变分不等式描述的控制系统(controlled system governed by variational inequality) 一类特定的由变分不等式描述的分布参数系统. 设 U,V 和 H 是希尔伯特空间，V 在 H 中稠密，$V\subset H$，V 嵌入 H 的映射是连续的，H 的对偶空间 $H'=H$，V 的对偶空间记为 V'，$V\subset H\subset V'$. 设 A 是 V 到 V'的线性有界算子，满足条件

$$\langle Av,v\rangle \geqslant \alpha\|v\|^2, \quad \forall v \in V,$$

这里$\langle\cdot,\cdot\rangle$是 V'和 V 的对偶积，$\|\cdot\|$是 V 的范数，$\alpha>0$ 是常数. 又设 $\varphi: V\to\mathbb{R}$ 是下半连续的凸函数，$\partial\varphi$是 φ 的次微分，通常 ∂p 是合系值映射 $f\in V'$，$B\in L(U,V')$. 由变分不等式

$$f + Bu \in Ay + \partial\varphi(y)$$

描述的受控系统，是非线性椭圆型控制系统. 当 $\varphi=0$ 时，它是线性椭圆型控制系统

$$Ay = f + Bu.$$

由变分不等式

$$f(t) + Bu(t) \in \dot{y}(t) = Ay(t) + \partial\varphi(y(t))$$

描述的受控系统，是非线性抛物型控制系统．当 $\varphi=0$ 时，它是线性抛物型控制系统

$$\dot{y}(t)+Ay(t)=f(t)+Bu(t),$$

这里 $f:\mathrm{R}\to V'$ 是连续的．

受控扩散过程(controlled diffusion process) 控制作用下的扩散过程．扩散是自然界中普遍存在的现象，比如气体的扩散、生物群的迁移以及肌肉注射药物后在身体内的扩散等．可用下述方程描述受控扩散过程

$$\frac{\partial C}{\partial t}=\operatorname{div}(D\operatorname{grad}C)+f(t,c,u), \qquad (1)$$

其中 C 是扩散物质的浓度，D 为扩散系数矩阵，grad 与 div 分别表示梯度与散度算子，u 是控制变量，它是空间和(或)时间变元的函数．u 变化在容许参数集 $\mathscr{U}_{ad}$ 内．任给 $u\in\mathscr{U}_{ad}$，可惟一确定浓度 C．因此扩散物质的浓度 C 受 u 的控制．模型(1)可用于描述半导体内的杂质扩散、生物群体的人为控制，以及工业开采石油后原油在油藏中渗透等现象．

分布参数系统的控制输入(control input of distributed parameter system) 针对分布参数系统特点施加控制作用的几种方式．对于 R^n 的某个区域 Ω 上由偏微分方程描述的控制系统，控制输入(亦称控制量)可以作用于 Ω 的一些子区域上，称为分布控制；也可以作用于 Ω 的边界 Γ 的一部分上，称为边界控制；也可以作用于区域或边界的某些点上，称为点控制．例如考察热传导系统

$$\begin{cases}\dfrac{\partial y}{\partial t}=\Delta y(x,t)+b(x)u(t) \ (x\in\Omega,t>0),\\ y|_{\Gamma}=0,\end{cases}$$

式中 Δ 为拉普拉斯算子，Ω 是 R^3 中的一个有界开区域，$\Gamma=\partial\Omega$ 是 Ω 的边界，$u(t)$ 是控制输入．当 $b(\cdot)\in L^p(\Omega)$，$1\leqslant p\leqslant\infty$ 时，就得到分布控制；当 $b(x)=\delta(x-b)$ 时，就是点控制的情况，这里 $\delta(x-b)$ 是作用于点 $b\in\Omega$ 的狄喇克 δ 函数．为了说明边界控制，仍以热传导系统为例，

$$\begin{cases}\dfrac{\partial y}{\partial t}=\Delta y(x,t)+f(x,t) \quad (x\in\Omega,t>0),\\ y(\xi,t)=u(\xi,t) \quad (\xi\in\Gamma,t>0),\end{cases}$$

这里，控制量 $u(\xi,t)$ 作用于 Ω 的边界 Γ 上，得到边界控制．如果 $u(\xi,t)=\delta(\xi-b)u(t)$，则得到边界上点控制的情况．

分布参数控制系统的能控性(controllability for distributed parameter control system) 一种表示分布参数控制系统控制能力的结构性质．在一个控制系统中，对哪些状态加以控制就能把系统引向预定的目标，这就是系统的能控性问题．由于分布参数控制系统的无穷维特点及其控制输入的多样性，比起集中参数控制系统，它的能控性问题要复杂得多．以如下的无穷维希尔伯特空间 X 中的线性控制系统为例加以说明：

$$\dot{x}(t)=Ax(t)+Bu(t). \qquad (1)$$

常用的能控性定义有精确能控性、近似能控性、精确零态能控性和近似零态能控性．

1．精确能控性．对于任意给定的系统初始态 $x(0)=x_0$ 和末态 x_1，都能找到一个与 x_0 和 x_1 无关的有穷时间 t_1 和一个控制 $u(t)(0\leqslant t\leqslant t_1)$，使得系统(1)在此控制 $u(t)$ 的作用下，从 $t=0$ 时刻的初态 x_0 出发，于 t_1 时刻达到状态 x_1，则称系统(1)在有穷时间区间 $[0,t_1]$ 上精确能控(有时也称为 (A,B) 在 $[0,t_1]$ 上精确能控，下同)．系统(1)在 $[0,t_1]$ 上精确能控的充分必要条件是存在常数 $\gamma>0$，使得

$$\left(\int_0^{t_1}\|B^*T^*(s)x^*\|^2\mathrm{d}s\right)^{\frac{1}{2}}\geqslant\gamma\|x^*\| \quad (\forall x^*\in X^*),$$

其中 X^* 为 X 的对偶空间，$T(t)$ 为由 A 生成的强连续线性算子半群，B^* 表示 B 的伴随．

2．近似能控性．称系统(1)在 $[0,t_1]$ 上近似能控，是指对于任意给定的初态 x_0 和末态 x_1，以及任意给定的正数 $\varepsilon>0$，都能找到一个控制 $u(t)(0\leqslant t\leqslant t_1)$，使得系统(1)在此控制 $u(t)$ 作用下从初态 x_0 出发，于 t_1 时刻达到的状态与 x_1 的距离不超过 ε，即

$$\left\|T(t_1)x_0+\int_0^{t_1}T(t_1-s)Bu(s)\mathrm{d}s-y_1\right\|<\varepsilon.$$

(A,B) 在 $[0,t_1]$ 上近似能控的充分必要条件是

$$x^*\in X^*,B^*T^*(t)x^*=0,\forall t\in[0,t_1]\Rightarrow x^*=0.$$

对于大多数分布参数控制系统来说，精确能控性的要求过高，故常使用近似能控性．

3．精确零态能控性．称系统(1)在有穷时间区间 $[0,t_1]$ 上精确零态能控，是指对于任意给定的初态 x_0，都能找到一个控制 $u(t)(0\leqslant t\leqslant t_1)$，使得系统(1)从 $t=0$ 时刻的初态 x_0 出发，在控制 $u(t)$ 作用下于 t_1 时刻达到静止态．

4．近似零态能控性．称系统(1)在有穷时间区间 $[0,t_1]$ 上近似零态能控，是指对于任意给定的初态 x_0 和任意正数 $\varepsilon>0$，都能找到一个控制 $u(t)(0\leqslant t\leqslant t_1)$，使得系统(1)从 $t=0$ 时刻的初态 x_0 出发，在控制 $u(t)$ 作用下于 t_1 时刻达到的状态与零态的距离不超过 ε，即

$$\left\|T(t_1)x_0+\int_0^{t_1}T(t_1-s)Bu(s)\mathrm{d}s\right\|<\varepsilon.$$

分布参数控制系统的能观测性(observability for distributed parameter control system) 一种表示分布参数控制系统中对状态观测能力的结构性质．对于一个系统的哪些量进行观测，就能够确定系统的全部状态，这是能观测性问题．分布参数控制系

统中常用的能观测性的定义有初态能观测性、连续初态能观测性、末态能观测性和连续末态能观测性.以如下具有输出(即观测)的无穷维线性系统

$$\begin{cases}\dot{z}(t) = Fz(t),\\ y(t) = Cz(t)\end{cases} \tag{1}$$

为例分别加以说明,这里 F 是希尔伯特空间 Z 中某个强连续线性算子半群 $S(t)$ 的无穷小生成算子,$y(t)$ 属于另一个希尔伯特空间 Y(观测空间或输出空间),C 是 Z 到 Y 的有界线性算子.

1.初态能观测性.称系统(1)在有穷时间区间 $[0,t_1]$ 上是初态能观测的(亦称 (F,C) 在 $[0,t_1]$ 上是初态能观测的,下同),是指通过系统的观测量 $y(t)(0\leqslant t\leqslant t_1)$,能惟一地确定系统的初态 $x(0)=x_0$. (F,C) 在 $[0,t_1]$ 上初态能观测的充分必要条件是

$$z\in Z,\ CS(t)z=0,\ \forall t\in[0,t_1]\Rightarrow z=0.$$

2.连续初态能观测性.称系统(1)在有穷时间区间 $[0,t_1]$ 上是连续初态能观测的,是指通过系统的观测量 $y(t)(0\leqslant t\leqslant t_1)$,能惟一地确定系统的初态 $x(0)=x_0$,并且当观测量出现小扰动时,所确定的初态也保持小偏差. (F,C) 在 $[0,t_1]$ 上连续初态能观测的充分必要条件是:存在常数 $\gamma>0$,使

$$\gamma\|CS(\cdot)z\|_{L^2(0,t_1;Y)}\geqslant\|z\|\quad(\forall z\in Z).$$

3.末态能观测性.称系统(1)在有穷时间区间 $[0,t_1]$ 上是末态能观测的,是指通过系统的观测量 $y(t)(0\leqslant t\leqslant t_1)$ 能惟一地确定系统在末时刻 t_1 的状态. (F,C) 在 $[0,t_1]$ 上末态能观测的充分必要条件是

$$z\in Z,\ CS(t)z=0,\ \forall t\in[0,t_1]\Rightarrow S(t_1)z=0.$$

4.连续末态能观测性.称系统(1)在有穷时间 $[0,t_1]$ 上是连续末态能观测的,是指通过系统的观测量 $y(t)(0\leqslant t\leqslant t_1)$,能惟一地确定系统的末时刻 t_1 的状态,并且当量测值出现小扰动时,所确定的末态也保持小偏差. (F,C) 在 $[0,t_1]$ 上连续末态能观测的充分必要条件是:存在常数 $\gamma>0$,使得

$$\gamma\|CS(\cdot)z\|_{L^2(0,t_1;Y)}\geqslant\|S(t_1)z\|\quad(\forall z\in Z).$$

能控性与能观测性的对偶关系(duality between controllability and observability) 无穷维线性控制系统的一种属性.与集中参数线性控制系统相类似,无穷维线性控制系统的能控性与能观测性之间有对偶关系,即一个系统的能控性正好对应于另一个相关系统的能观测性.具体地说,令 $U=Y^*$, $X=Z^*$, $A=F^*$, $B=C^*$, $T(t)=S(t)^*$,其中空间的 * 表示对偶,算子的 * 表示伴随,并且 X,U,Y,Z 均为自反巴拿赫空间,那么:

1. (F,C) 在 $[0,t_1]$ 上初态能观测,当且仅当 (A,B) 在 $[0,t_1]$ 上近似能控.

2. (F,C) 在 $[0,t_1]$ 上连续初态能观测,当且仅当 (A,B) 在 $[0,t_1]$ 上精确能控.

3. (F,C) 在 $[0,t_1]$ 上末态能观测,当且仅当 (A,B) 在 $[0,t_1]$ 上近似零态能控.

4. (F,C) 在 $[0,t_1]$ 上连续末态能观测,当且仅当 (A,B) 在 $[0,t_1]$ 上精确零态能控.

无穷维线性系统的稳定性(stability of infinite dimensional linear system) 无穷维线性系统的一种结构性质.对于巴拿赫空间 X 中的线性系统

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t} = Ax(t),\\ x(0) = x_0,\end{cases} \tag{1}$$

其中 A 为 X 中某个强连续算子半群 $T(t)$ 的无穷小生成算子,稳定性的定义基本上有如下四种:

1.称系统(1)为指数渐近稳定的,简称指数稳定的,是指存在常数 $M>0$ 和 $\omega<0$,使得

$$\|T(t)\|\leqslant Me^{\omega t}\quad(\forall t\geqslant 0).$$

2.称系统(1)为强渐近稳定的,简称强稳定的,是指 $T(t)$ 满足

$$\lim_{t\to\infty}\|T(t)x\| = 0\quad(\forall x\in X).$$

3.称系统(1)为弱渐近稳定的,简称弱稳定的,是指

$$\lim_{t\to\infty}\langle T(t)x,x^*\rangle=0\quad(\forall x\in X,\forall x^*\in X^*).$$

4.称系统(1)为一致稳定的,亦称一致有界的,是指存在常数 $M>0$,使得 $\|T(t)\|\leqslant M,\forall t\geqslant 0$.

系统(1)的稳定性有时也称为半群 $T(t)$ 的稳定性.

谱确定增长假设(spectrum determined growth assumption) 巴拿赫空间中线性控制系统稳定性的一类重要条件.巴拿赫空间 X 中以 A 为生成算子的强连续半群 $T(t)$ 称为满足谱确定增长假设,是指

$$s(A)\triangleq\sup\{\mathrm{Re}\,\lambda\,|\,\lambda\in\sigma(A)\}$$
$$=\lim_{t\to\infty}\frac{\lg\|T(t)\|}{t}\triangleq\omega(A),$$

$s(A)$ 通常称为 A 的谱界,而 $\omega(A)$ 则称为由 A 所生成的半群 $T(t)$ 的增长界.于是在谱确定增长假设成立之下,线性系统

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t} = Ax(t),\\ x(0) = x_0\end{cases}$$

的稳定性完全由 A 的谱决定.特别当 $s(A)<0$ 时,上述系统指数稳定.已经知道在下列条件之一满足时,谱确定增长假设成立:

1. A 为有界线性算子.

2. $T(t)$ 对 $t>t_0$($t_0\geqslant 0$ 为某固定数)为可微半群.

3. $T(t)$ 对 $t>t_0$($t_0\geqslant 0$ 为某固定数)为紧半群.

4.存在某个数 $\xi>0$,使得 $r(T(\xi))=\sup\{|\lambda|\,|\,\lambda\in\sigma_p(T(\xi))\cup\sigma_r(T(\xi))\cup\{0\}\}$,其中 $r(T)$ 表示有界线性算子 T 的谱半径,$\sigma_p(T)$ 和 $\sigma_r(T)$ 分别表示线性

算子 T 的点谱和剩余谱.

5. $T(t)$ 为 $L^p(\Omega)$ 上正半群，$1\leqslant p\leqslant\infty$，这里 $\Omega\subset\mathbf{R}^n$ 为可测子集，$L^p(\Omega)$ 的正算子 S 指 $S\varphi\in K$，$\forall\varphi\in K\triangleq\{\varphi\in L^p(\Omega)\mid\varphi(x)\geqslant0,\text{a.e. } x\in\Omega\}$，这里 a. e. 表示几乎处处的意思.

分布参数系统的最优控制(optimal control for distributed parameter system)　无穷维系统关于性能指标取极小(或极大)的控制问题. 它是高维空间中变分问题的现代发展. 对于分布参数系统的最优控制，有例表明庞特里亚金极大值原理可能不成立，但对于巴拿赫空间 X 中发展方程

$$x(t)=G(t,0)x_0+\int_0^t G(t,s)f(s,x(s),u(s))\mathrm{d}s$$

$$(0\leqslant t\leqslant 1,\ x(1)\in Q\subset X,\ u(t)\in U)$$

描述的分布参数系统，关于性能指标

$$J(u(\cdot))=\int_0^1 g(t,x(t),u(t))\mathrm{d}t$$

取极小的最优控制问题，却有：假设 G,f,g 满足适当条件，U 是距离空间中的一给定集合，那么，当 Q 在 X 中具有限余维数(即 X 可以分解成 Q 和另一个有限维子空间的直接和)时，最优控制满足极大值原理.

分布参数系统的时间最优控制(time optimal control for distributed parameter system)　以时间作为性能指标的一种最优控制问题. 分布参数系统在控制作用受约束时，寻求把系统从给定状态 x_0 转移到 x_1，使时间最短的控制问题，称为时间最优控制. 对于无穷维线性系统，等时区可能不凸，但等时区的闭包是凸的，利用它可以讨论线性系统的时间最优控制.

无穷维线性系统的二次最优控制(quadratic optimal control for infinite dimensional linear system)　以状态变量和控制变量的二次型作为性能指标的一种最优控制问题. 无穷维希尔伯特空间中的线性系统

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=Ax(t)+Bu(t)\quad(0<t<T),$$

$$x(0)=x_0,$$

关于二次性能指标

$$J(u(\cdot))=\int_0^T\{\langle Qx(t),x(t)\rangle+\langle Ru(t),u(t)\rangle\}\mathrm{d}t+\langle Q_1x(T),x(T)\rangle$$

取最小的控制问题，是无穷维线性系统的二次最优控制问题. 这里 A 是 C_0 算子半群的母元，B 是线性有界算子，$Q^*=Q$ 和 $Q_1^*=Q_1$ 是非负的，$R^*=R\geqslant\delta I>0$，其中 δ 为一正数，上标 * 表示矩阵的复共轭转置，$\langle\cdot,\cdot\rangle$ 是相应空间中的内积. 该问题可以转化为希尔伯特空间中二次泛函的极值问题，并且导出最优控制是状态反馈

$$u(t)=-R^{-1}B^*P(t)x(t),$$

其中 $P^*(\cdot)=P(\cdot)$ 满足无穷维黎卡提微分方程. 对于由椭圆型偏微分方程、抛物型偏微分方程、双曲型偏微分方程、线性泛函微分方程描述的分布参数系统，关于二次性能指标的最优控制问题有许多具体结果，当控制作用是点作用或边界作用时，算子 B 将是无界的. 无穷维线性系统的二次最优控制问题仍在深入讨论中.

无穷维黎卡提微分方程(Riccati differential equation for infinite dimensional system)　无穷维线性系统二次最优控制问题中引出的一类算子微分方程. 由无穷维线性系统二次最优控制问题引入的算子微分方程

$$-\frac{\mathrm{d}P(t)}{\mathrm{d}t}=Q+P(t)A+A^*P(t)-P(t)MP(t)\quad(0\leqslant t\leqslant T),$$

$$P(T)=Q_1$$

称为无穷维黎卡提微分方程，这里 A 是 C_0 算子半群的母元，$Q^*=Q\geqslant0$，$Q_1^*=Q_1\geqslant0$，$R^*=R\geqslant\delta I>0$，$M=BR^{-1}B^*$，B 是线性有界算子，而 $P^*(t)=P(t)$ 是无穷维希尔伯特空间上的对称算子. 该黎卡提微分方程的解 $P(\cdot)$，可以视为下述内积黎卡提微分方程

$$-\frac{\mathrm{d}}{\mathrm{d}t}\langle P(t)x,y\rangle=\langle Qx,y\rangle+\langle P(t)Ax,y\rangle+\langle P(t)x,Ay\rangle-\langle MP(t)x,P(t)y\rangle,\ \langle P(T)x,y\rangle=\langle Q_1x,y\rangle$$

的解，其中 $x,y\in D(A)$. 也可以把 $P(\cdot)$ 视为黎卡提积分方程

$$P(t)=\mathrm{e}^{A^*(T-t)}Q_1\mathrm{e}^{A(T-t)}+\int_t^T\mathrm{e}^{A^*(s-t)}[Q-P(t)MP(t)]\mathrm{e}^{A(s-t)}\mathrm{d}s$$

的解.

无穷维黎卡提代数方程(Riccati algebraic equations for infinite dimensional system)　确定无穷维线性系统二次最优控制问题的一种特殊黎卡提方程. 无穷维线性系统为

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=Ax(t)+Bu(t)\quad(0\leqslant t<\infty),$$

$$x(0)=x_0,$$

关于无限时间的二次性能指标

$$J(u(\cdot))=\int_0^\infty\{\langle Qx(t),x(t)\rangle+\langle Ru(t),u(t)\rangle\}\mathrm{d}t$$

取最小的最优控制问题研究中，引入黎卡提算子方程

$$Q+PA+A^*P-PMP=0,\tag{1}$$

其中 A 是 C_0 算子半群的母元，B 是线性有界算子，

$M=BR^{-1}B^*, Q^*=Q\geqslant 0, R^*=R\geqslant \delta I>0$. 黎卡提代数方程(1)的解可视为内积形式的黎卡提方程

$$\langle Qx,y\rangle+\langle PAx,y\rangle+\langle Px,Ay\rangle-\langle MPx,Py\rangle=0$$

的解,其中 $x,y\in D(A)$;也可视 P 适合积分等式

$$P=\int_0^\infty e^{A^*t}[Q-PMP]e^{At}dt.$$

分布参数系统的反馈镇定(feedback stabilization of distributed parameter system) 运用反馈使分布参数系统达到稳定的一种重要手段.通常分为状态反馈镇定和输出反馈镇定.一般说来,能够实现的控制器输入和传感器输出都是有穷维的,因此在应用状态反馈时必须先设计观测器.常用的处理办法是采用振型分析法或有限元法把分布参数系统化成近似的集中参数系统,然后再应用有穷维控制系统理论设计反馈控制器.

分布参数系统的能稳性(stabilizability of distributed parameter system) 分布参数系统能通过反馈控制达到稳定的一种结构性质.稳定性是控制系统要达到的最重要的目标之一.考虑巴拿赫空间 X 中线性控制系统

$$\begin{cases}\dot{x}(t)=Ax(t)+Bu(t),\\ x(0)=x_0,\end{cases}$$

其中 A 生成 X 中强连续线性算子半群 $T(t)$,$u\in U$(控制空间).如果存在一个反馈(增益)算子 $F:U\to X$,使得 $A-BF$ 产生指数稳定的(强稳定的、弱稳定的或一致稳定的)强连续线性算子半群,则系统(或称偶对(A,B))是指数能稳的(强能稳的、弱能稳的或一致能稳的).

分布参数系统的能检测性(detectability of distributed parameter system) 在某种意义上与能稳性相对偶的一种概念.考虑巴拿赫空间 X 中附加输出方程的线性系统

$$\begin{cases}\dot{x}(t)=Ax(t)+Bu,\\ y(t)=Cx(t),\end{cases}$$

其中 A 生成 X 中强连续线性算子半群,$C:X\to Y$(Y 为观测空间)为有界线性算子.称系统(或偶对(A,C))是指数能检测的(强能检测的、弱能检测的或一致能检测的),是指存在一个线性算子 $K:Y\to X$,使得 $A-KC$ 产生 X 上指数稳定的(强稳定的、弱稳定的或一致稳定的)强连续线性算子半群.

模态分析法(mode analysis method) 亦称振型分析法.对分布参数系统进行分析和控制器设计的一种近似方法.在研究分布参数系统,特别是弹性振动系统时,系统主算子的各个本征值对应于系统的各阶固有振动频率,而相应的本征函数则称为系统的相应阶的振型或模态.例如,考察弹性振动方程

$$\frac{d^2y(t)}{dt^2}+Ay(t)+B\frac{dy(t)}{dt}=0, \qquad (1)$$

其中表示系统振动的量 $y(t)$ 属于某个适当的函数空间 H(状态空间),A 是振动算子,它在 H 中是正定自伴的,B 表示系统内阻尼算子.设$\{\lambda_n|n\geqslant 1\}$为 A 的本征值集,$\{\varphi_n|n\geqslant 1\}$为相应的本征函数集,于是 $\omega_n=\sqrt{\lambda_n}$ 为系统 n 阶固有振动频率,而 φ_n 称为系统的 n 阶振型.通常情况下,系统(1)的振动 $y(t)$ 可以展成$\{\varphi_n\}$的无穷级数(傅里叶级数):

$$y(t)=\sum_{n=1}^{\infty}q_n(t)\varphi_n.$$

在工程实践中,由于阻尼的作用,往往只有前几个振型起主导作用,高阶振型随着时间很快消失.因此,在对系统进行分析和控制器设计中,大多仅用有穷个振型展开的近似方法,即

$$y(t)=\sum_{n=1}^{N}q_n(t)\varphi_n.$$

然后在此基础上进行系统分析和控制器设计,这就是所谓的振型分析法.

控制溢出(control spillover) 由于无穷维系统用有穷维系统逼近所引起的控制信号对残差系统的作用.在工程实践中,无穷维系统常用有穷维近似模型代替,并基于这种有穷维模型设计反馈控制器.例如,考虑无穷维希尔伯特空间 H 中的线性系统

$$\begin{cases}\dfrac{dx(t)}{dt}=Ax(t)+Bu(t),x(0)=x_0,\\ y(t)=Cx(t).\end{cases} \qquad (1)$$

设 H 能分解成两个子空间 H_N 和 H_R 的正交和:$H=H_N\oplus H_R$,其中 H_N 为有穷维子空间,并且 H_N 和 H_R 均为 A 的约化子空间,于是无穷维系统(1)可分解成

$$\dot{x}_N(t)=A_Nx_N(t)+B_Nu(t),\ x_N(0)=P_Nx_0, \qquad (2)$$

$$\dot{x}_R(t)=A_Rx_R(t)+B_Ru(t),\ x_R(0)=P_Rx_0, \qquad (3)$$

$$y(t)=C_Nx_N(t)+C_Rx_R(t), \qquad (4)$$

其中 $x_N=P_Nx, x_R=P_Rx, A_N=P_NAP_N, B_N=P_NB, C_N=CP_N, C_R=CP_R$ 等,并且 P_N 和 P_R 为相应子空间上的直交投影算子.对应于(1)的降阶模型为

$$\begin{cases}\dot{x}_N(t)=A_Nx_N(t)+B_Nu(t),\ x_N(0)=P_Nx_0,\\ \tilde{y}(t)=C_Nx_N(t).\end{cases} \qquad (5)$$

这里的残差子系统(3)中出现的 $B_Ru(t)$ 项称为控制溢出,而(4)中出现的 $C_Rx_R(t)$ 项则称为观测溢出.通常人们基于有穷维模型(5)设计反馈控制器.显然,反馈控制器正是通过控制溢出和观测溢出进入残差方程(4)而发生作用.一般地,由于控制溢出和观测溢出的存在,基于有穷维近似模型(5)设计的反馈控制器未必能保证无穷维系统(1)稳定.因此在设计控制器时必须考虑控制溢出和观测溢出的影响.

观测溢出(observation spillover) 由于无穷维系统用有穷维系统近似所引起的残差系统观测信号

（参见“控制溢出”）.

数学物理反问题（inverse problem of mathematical physics） 根据实测数据确定数学物理方程系数（或参数）的一种建模问题. 人们通过对自然现象、工业过程（所谓系统）的深入研究，将系统的输入信号 x 与输出信号 y 建立联系：$\Phi(x,q)=y$，其中 q 是参数. 当 q 已知时，给出 x 便可求出 y，这就是数学物理原问题；当 q 未知时，必须通过 x 与 y 的量测值，反求出 q，这就是数学物理反问题. 当模型 Φ 是用偏微分方程的定解问题描述时，q 可以是方程中的未知系数、边值和（或）始值条件，这就是偏微分（或数学物理）方程反问题.

分布参数系统的状态估计（state estimation for distributed parameter system） 从含有噪声的量测数据出发估计随机分布参数系统状态向量的方法. 考虑随机分布参数系统 $\Phi(t,y,w)=0$，其中 y 是状态向量，w 取值于函数空间的维纳过程. 对于这类模型，只能利用量测 $z=\Lambda y+\xi$，并用统计方法来估计状态向量 y，此处 Λ 是已知的量测算子，ξ 是取值于某函数空间的维纳过程. 这就是分布参数系统的状态估计. 依据被估计状态与所用数据间的相对关系，状态估计分为平滑、滤波和预报. 再根据所用数学工具的不同，估计法分为“严格法”与“形式法”. 这两类方法都可将集中参数系统的卡尔曼-布西滤波理论推广到随机分布参数系统.

分布参数系统的辨识（identification of distributed parameter system） 根据实测数据确定分布参数系统模型的一种方法. 先通过机理分析，可以导出分布参数系统模型

$$\Phi(y,q)=0 \tag{1}$$

其中 $y\in V$ 是刻画系统特性的状态向量，$q\in Q_{ad}$ 是待定参数，Q_{ad} 是容许参数集，然后依据系统的观测输出 $z=\Lambda y$（Λ 是已知的量测算子），应用数学方法确定模型（1）的参数 $\hat{q}\in Q_{ad}$，使得模型之输出依某种方式与 z 相（或尽可能相）符合. 若模型（1）代表一个微分方程的定解问题，参数 q 可以是方程的系数、边值或始值，此时的参数辨识问题就是微分方程的反问题；当未知参数 q 是方程定义区域的边界时，确定 q 的问题就是几何形状设计问题，模型（1）也可解释为随机分布参数系统. 当模型（1）理解为随机分布参数系统时，确定参数 q 的问题就是统计学上的参数估计问题，只是此处的参数可以是一个函数. 任取 $q\in Q_{ad}$，假定自（1）可以惟一确定 y；且 y 关于 q 是稳定的（即 y 连续依赖于 q）. 为了表示这种依赖关系，记 $y=y(q)$. 设能依据量测输出 $z=\Lambda y$（Λ 是已知的量测算子）经用某种方法确定模型（1）的参数为 $\hat{q}\in Q_{ad}$，则考虑误差 $e=z-\Lambda y(\hat{q})$，倘若由 $e=0$，便可得到 $\hat{q}$ 就是（1）的真实参数，因而 $\hat{q}$ 惟一地被确定出，则称模型（1）是能辨识的. 当（1）是一个微分方程的定解问题时，参数的能辨识性就是微分方程反问题解的惟一性问题.

最优控制理论

最优控制理论（optimal control theory） 现代控制理论中最早发展的分支之一. 对于受控系统，常常要求找到控制，使得在这控制作用下，系统从一个状态转移到所希望的状态，并且还要求系统的某种性能最好，这称为最优控制问题. 最优控制理论乃是研究最优控制满足的条件、最优控制的存在性及惟一性、最优控制问题求解的理论和方法. 当受控对象的动态由常微分方程（差分方程）描述时，有集中（总）参数系统的最优控制理论；当受控对象的动态由偏微分方程（积分方程）描述时，有分布参数系统的最优控制理论. 同样，有随机系统的最优控制理论. 20 世纪 50 年代和 60 年代中期，贝尔曼（Bellman, R.）的最优性原理、庞特里亚金（Понтрягин, Л. С.）的极大值原理和线性时间最优控制理论以及卡尔曼（Kalman, R. E.）的线性二次最优控制理论的出现标志着确定性集中参数系统最优控制理论的形成. 往后，一方面，集中参数系统最优控制理论的存在性、求解方法和算法以及在多方面的应用等不断发展；另一方面，分布参数控制系统、随机控制系统以及非工程系统的最优控制理论也相继提出并形成. 本节仅以连续时间集中参数系统为例，叙述与最优控制理论有关的若干问题. 设有受控对象，它的动力学由下列向量常微分方程

$$\dot{x}=f(t,x,u) \tag{1}$$

描述，简称受控系统. 其中 x 表示受控系统的状态向量，它属于 n 维欧几里得空间，记为 $x\in\mathbb{R}^n$；u 表示受控系统的控制向量，通常不能任意取值而受到限制，用属于 r 维欧几里得空间中某闭集 U_r 表示 u 所受的限制，记为 $u\in U_r\subset\mathbb{R}^r$. 通常有 $r\leqslant n$. 设 x_0 是系统（1）的控制作用开始时刻的状态，即初始状态，x_f 是控制作用结束时刻的希望状态，即终端状态. x_0，x_f 可以是自由的，亦可是事先给定的或受到限制的. 用定义在某函数集合 $\mathscr{U}$ 上的泛函

$$J[u(\cdot)]=\int_{t_0}^{t_f}f^0(t,x(t),u(t))\mathrm{d}t \tag{2}$$

来刻画系统（1）的性能的度量，称为性能指标. 式（2）中 t_0，t_f 分别是控制作用开始和结束的时刻，它们可以被事先给定，亦可待求；f^0 是给定的、变量 t,x,u 的连续（或可积）标量函数. 对于 $u(t)\in\mathscr{U}$，用 $x(t)$ 表示向量微分方程 $\dot{x}=f(t,x,u(t))$ 的满足初始条件 $x(t_0)=x_0$ 和终端条件 $x(t_f)=x_f$ 的解，简称系统（1）的对应于 $u(t)$ 的解，或称 $x(t)$ 是系统（1）在控制

$u(t)$作用下的轨线.

受控过程(controlled process) 人们对系统施加影响而使之达到一定目的的过程.它体现了人的主观能动性在改造世界中的作用.在数学上,这种过程通常由微分方程(差分方程、积分方程等)来描述.集中参数控制过程由常微分方程(组)描述,分布参数控制过程由偏微分方程描述等.例如,对于集中参数控制系统,当用$x(t)$表示系统的n维状态向量函数,$u(t)$表示r维控制向量函数时,该受控过程可用下面常微分方程描述$\dot{x}=f(t,x,u)$,其中f是定义在$[t_0,t_f]\times\mathrm{R}^n\times\mathrm{R}^r$中的、并满足一定条件的$n$维向量函数.

容许控制(admissible control) 满足限制条件并使所讨论受控系统的初值和终端值问题有解的控制函数.定义在时间区间$[t_0,t_f]$上的控制函数$u(t)$称为容许控制,如果它的取值满足限制条件并使下面方程

$$\dot{x}=f(t,x,u(t)) \tag{1}$$

同时满足初始条件$x(t_0)=x_0$和终端条件$x(t_f)=x_f$的解$x(t)$在$[t_0,t_f]$上存在惟一.容许控制全体组成的集合称为容许控制函数集合,用$\mathscr{U}_{ad}$表示.容许控制函数集合由对最优控制的要求来规定.例如,可取定义在区间$[t_0,t_f]$上、取值于U^r内,并使系统(1)的解满足初始条件和终端条件的分段连续向量值函数全体为容许控制函数的集合.

控制域(control domain) 对控制系统中控制器作用大小的限制.控制系统中控制器都是由具体的物理装置构成,其控制作用的大小不是随意的,而是取决于物理装置的性能.这种控制作用在数学上可描述为控制量的量值要限定于空间的某一个合适的集合(闭集、有界集、多面体等)上,也可以由控制量的某个函数在时间域上的积分来限定.例如,若某系统的控制量是r维向量函数,记为

$$u(t)=[u_1(t),u_2(t),\cdots,u_r(t)]^{\mathrm{T}},$$

其中每个分量最大幅值都是1,则该控制系统的控制域U_r即是r维欧氏空间中对称于原点立方体,其数学表达式为

$$U=\{(u_1,u_2,\cdots,u_r):|u_i|\leqslant 1,i=1,2,\cdots,r\}.$$

有界控制(bounded control) 值域属于欧氏空间中的有界集合的控制函数.记r维欧氏空间R^r中的集合U是系统的控制域.称定义在时间区间$[t_0,t_f]$上的控制函数$u(t)$是有界控制,如果R^r中的集合$\{\xi:\xi=u(t),t\in[t_0,t_f]\}$的闭包是有界集.

终端条件(terminal condition) 控制系统在控制作用下终端状态满足的条件.例如,最终状态是状态空间中一个点,最终状态落在状态空间中的一条曲线、一个曲面上或属于一个集合等.最终状态所在的流形称为终值流形.

最优控制(optimal control) 使性能指标达到极小(或极大)的容许控制.对于"最优控制理论"条目中的系统(1)和性能指标(2),其最优控制问题指在容许控制集合$\mathscr{U}_{ad}$中寻找一控制,它使(2)中$J[u(\cdot)]$达到极小(或极大).使$J[u(\cdot)]$达到极小(或极大)的容许控制称为(1),(2)的最优控制,记为$u^*(t)$,(1)的对应于$u^*(t)$的解$x^*(t)$称为(1),(2)的最优轨线,对应于$u^*(t)$,$x^*(t)$的性能指标值

$$J^*\triangleq J[u(\cdot)]=\int_{t_0}^{t_f^*}f^0(t,x^*(t),u^*(t))\mathrm{d}t$$

称为最优指标值,其中t_f^*是最优控制$u^*(t)$作用结束的时刻(终端时刻t_f为不固定的情况),称为最优时刻.

最优轨线(optimal trajectory) 受控系统的对应于最优控制的轨线.见"最优控制".

极大值原理(maximum principle) 最优控制应满足的必要条件.是庞特里亚金(Понтрягин,Л.С.)等人于1958年提出的,全称是庞特里亚金极大值原理.这里就下述简单情况进行叙述:若以下方程(1),(2)中f^0,f不显含时间t;终端状态x_f受形如$g(x_f)=0$的约束,$g(x)$是x的连续可微q维向量函数($q\leqslant n$),即以

$$\dot{x}=f(x,u), \tag{1}$$

$$J[u(\cdot)]=\int_{t_0}^{t_f}f^0(x(t),u(t))\mathrm{d}t, \tag{2}$$

$$g(x_f)=0 \tag{3}$$

为例叙述极大值原理.定义(1),(2)的哈密顿函数$H(x,\psi_0,\psi,u)$为

$$H(x,\psi_0,\psi,u)\triangleq\psi_0 f^0+\psi^{\mathrm{T}}f(x,u).$$

并记

$$\mathscr{M}(x,\psi_0,\psi)\triangleq\sup_{u\in U}H(x,\psi_0,\psi,u).$$

极大值原理叙述为:如果$u^*(t)$是系统(1)和性能指标(2)的最优控制,$x^*(t)$是相应的最优轨线,那么必定存在非零函数$\psi_0(t)$,$\psi(t)$,它们和$u^*(t)$,$x^*(t)$一起在区间$[t_0,t_f]$上满足下列条件:

1. $\dot{x}^*(t)=f(x^*(t),u^*(t))$, $x^*(t_0)=x_0$, (4)

$$\dot{\psi}(t)=-\left(\frac{\partial f}{\partial x}\right)^{\mathrm{T}}_{\substack{x=x^*(t)\\u=u^*(t)}}\cdot\psi(t)-\left(\frac{\partial f^0}{\partial x}\right)^{\mathrm{T}}_{\substack{x=x^*(t)\\u=u^*(t)}}\cdot\psi_0(t), \tag{5$_1$}$$

$$\dot{\psi}_0(t)=0. \tag{5$_2$}$$

2. 在$[t_0,t_f]$上除有限多个时刻外处处成立:

$$\mathscr{M}(x^*(t),\psi_0(t),\psi(t))=H(x^*(t),\psi_0(t),\psi(t),u^*(t)).$$

3. 当$t=t_f^*$时,有

$\psi_0(t_f^*)\leqslant 0$,$\mathscr{M}(x^*(t_f^*),\psi_0(t_f^*),\psi(t_f^*))=0$,

$$\psi(t_f^*)=\left(\frac{\partial g}{\partial x}\right)^{\mathrm T}_{x=x^*(t_f^*)}\cdot\mu,\qquad(6)$$

μ 为待定 q 维常向量.

又若上述条件 1 和条件 2 成立,则 $\psi_0(t)$, $\mathscr{M}(x^*(t),\psi_0(t),\psi(t))$ 在区间 $[t_0,t_f]$ 上皆恒等于常数. 上面诸式中的 ψ_0,ψ 称为(1),(2)的共轭变量或协态. 协态满足的微分方程 $(5)_1$,$(5)_2$ 称为共轭方程,$\psi(t_f^*)$满足的终端条件(6)称为横截条件. 条件 2 称为极大条件. 利用哈密顿函数的定义能将(4),$(5)_1$,$(5)_2$ 及(6)表示成如下形式:

$$\dot{x}^*(t)=\frac{\partial H}{\partial \psi^{\mathrm T}}\bigg|_{\substack{x=x^*(t)\\ \psi_0=\psi_0(t)\\ \psi=\psi(t)\\ u=u^*(t)}},\ x^*(t_0)=x_0;$$

$$\dot{\psi}_0(t)=0,\ \psi_0(t_f^*)\leqslant 0;$$

$$\dot{\psi}(t)=-\left(\frac{\partial H}{\partial x}\right)^{\mathrm T}\bigg|_{\substack{x=x^*(t)\\ \psi_0=\psi_0(t)\\ \psi=\psi(t)\\ u=u^*(t)}};$$

$$\psi(t_f)=\left(\frac{\partial g}{\partial x}\right)^{\mathrm T}_{x=x^*(t_f^*)}\cdot\mu.$$

当(1),(2)中 f,f^0 显含时间 t 时,也有相应的极大值原理. 极大值原理包含了为确定 $u(t),x(t)$, $\psi_0(t),\psi(t)$的全部关系式,但要具体求得这些函数并不容易. 在某些特殊情况下,例如(1)中 $f(x,u)$是 x 和 u 的线性函数,$J[u(\cdot)]$是工程上常用到的性能指标,这时从极大值原理能够求得其最优控制和最优轨线. 在一般情况下,可采用数学与计算机结合的办法求解.

最优控制算法(optimal control algorithm) 用数学和计算机相结合求解最优控制问题的方法. 有两类最优控制算法:一类称为直接算法;另一类称为间接算法. 直接算法由直接从系统的状态方程和性能指标出发,通过包括:容许控制参数化、采用以状态方程为约束关系的规划方法参数化、由参数构造相应控制函数、得到控制函数和相应轨线以组成最优控制问题的"近似解"等在内的一系列算法组成. 近似解的含义根据算法中采用的数学方法及最优控制问题的精度要求而定;间接算法从极大值原理的一组关系式出发,通过求解"两点边界值问题"的一系列算法,包括:求使哈密顿函数 $H(x,\psi_0,\psi,u)$关于 u 取极大的算法而得到 $u(x,\psi_0,\psi)$,将 $u(x,\psi_0,\psi)$代入状态方程和共轭方程,利用初始条件和横截条件求解两点边界值问题的算法,如梯度法、不变嵌入法和微分动态规划法等,以得到 $x(t),\psi_0(t),\psi(t)$(通常不止一组),将

$$x(t),\ u(t)\triangleq u(x(t),\psi_0(t),\psi(t))$$

代入性能指标进行比较的算法等. 相应于性能指标最小的那组 $x(t),u(t)$,在最优控制存在的条件下,即组成所论最优控制问题的解.

最优控制罚函数方法(penalty function method for optimal control) 求解状态和控制受到限制时的最优控制问题的一种计算方法. 在"最优控制"和"最优轨线"条目中,取容许控制函数集合为有界可测函数的等价类空间 $L^\infty\triangleq L^\infty([t_0,t_1],\mathrm{R}^m)$,赋 $u(\cdot)$以范数,L^∞是巴拿赫空间;则最优控制问题等价于(不受限)最优化问题:求解

(UP) $\inf\{J(u(\cdot))\,|\,u(\cdot)\in L^\infty([t_0,t_1],\mathrm{R}^m)\}$. (1)

实际问题中,控制和状态通常都受限. 如 $g(x(t_1))=0,s(t,x(t),u(t))\leqslant 0,t\in[t_0,t_1]$等,这里 $g:\mathrm{R}^n\to\mathrm{R}^\alpha$, $s:\mathrm{R}\times\mathrm{R}^n\times\mathrm{R}^m\to\mathrm{R}^\beta$ 为连续函数,α,β 为正整数;容许控制集 $B_1=\{u(\cdot)\,|\,u(\cdot)\in L^\infty,g(x(t_1,u(\cdot)))=0\}$;$B_2=\{u(\cdot)\,|\,u(\cdot)\in L^\infty,s(x(t,u(\cdot)),u(t))\leqslant 0,t\in[t_0,t_1]\}$均为 L^∞内的闭集. 这样,若最优控制问题中其状态和控制还受到限制,则这种受限最优控制问题可转化等价于受限最优化问题:求解

(CP) $\inf\{J(u(\cdot))\,|\,u(\cdot)\in L^\infty,u(\cdot)\in B\}$, (2)

B 是 L^∞内的闭集.

最常见的罚函数有内罚函数和外罚函数. 两者只是在受限集 B 的内部 $\mathring{B}$、边界 ∂B 及外部 $L^\infty\backslash B$ 表现有所不同而已. 下面统一地讲述内罚函数和外罚函数方法,在出现括号时,紧靠括号前的和括号内的符号、文字分别对应于内罚函数和外罚函数,其余则两者共同使用.

设 $\mathring{B}(B)$非空,凡满足下列三个性质的函数列 $\{P_i(u)\}$称为关于集 B 的内(外)罚函数列:

1. 在 $\mathring{B}(L^\infty)$内连续非负.

2. 对 $\mathring{B}(B)$内任一点 u 均有 $\overline{\lim\limits_{i\to\infty}}P_i(u)=0$.

3. 对任给定的点列$\{u_j\}\subset\mathring{B}(L^\infty\backslash B)$,$u_j\to u_*\in\partial B(L^\infty/B)$,均有

$$\lim_{j\to+\infty}P_i(u_j)\left[\lim_{i,j\to+\infty}P_i(u_j)\right]=+\infty.$$

例如,

$$P_i'(u)=-\varepsilon_i\int_{t_0}^{t_1}\sum_{j=1}^{\beta}\lg\left(\frac{-s^j(t,x(t,u),u(t))}{M}\right)\mathrm{d}t$$

和

$$P_i''(u)=\frac{1}{\varepsilon_i}\int_{t_0}^{t_1}\sum_{j=1}^{\beta}(\max\{0,s^j(t,x(t,u),u(t))\})^2\mathrm{d}t$$

分别是前述控制集 B_2 的内罚函数和外罚函数,此外 $\varepsilon_i>0,\varepsilon_i\to 0,-M$ 是 $s(t,x,u)$的下界,$M>0$.

现考虑不受限最优化问题序列:

$(\mathrm{UPI})_i$ $\inf\{J_i(u)\,|\,u\in\mathring{B}\},i=1,2,\cdots$, (3)

$(\mathrm{UPE})_i$ $\inf\{J_i(u)\,|\,u\in L^\infty\},i=1,2,\cdots$, (4)

其中 $J_i(u)=J(u)+P_i(u)$. 在一定的条件下可得:带罚函数的不受限最优化问题$(\mathrm{UPI})_i[(\mathrm{UPE})_i]$,$i=1,2,\cdots$,其解序列$\{u_i\}$的任一聚点必是受限最优化问题(CP)的解. 这样就把求解受限最优控制问题

转化为求解不受限的最优控制问题序列了，这是最优控制罚函数方法的基本思想.

极值控制系统(extreme control system)　由极大条件得到的控制函数带入哈密顿系统后形成的系统. 由哈密顿函数 $H(x,\psi_0,\psi,u)$ 构成如下系统：

$$\begin{cases}\dot{x}=\dfrac{\partial H(x,\psi_0,\psi,u)}{\partial\psi^{\mathrm{T}}}=f(x,u),\\ \dot{x}_0=f^0(x,u),\\ \dot{\psi}=-\left(\dfrac{\partial H(x,\psi_0,\psi,u)}{\partial x}\right)^{\mathrm{T}}\\ \qquad-\left(\dfrac{\partial f^0(x,u)}{\partial x}\right)^{\mathrm{T}}\psi_0,\\ \dot{\psi}_0=0,\end{cases}$$

称此为"极大值原理"条目中(1),(2)的哈密顿系统. 取

$$x(t_0)=x_0,\ \psi_0(t_0)\leqslant 0,$$

$$\psi(t_f)=\left(\frac{\partial g(x(t_f))}{\partial x(t_f)}\right)^{\mathrm{T}}\mu$$

作为哈密顿系统的端点条件. 记使 $H(x,\psi_0,\psi,u)$ 关于 u 达到极大的 u 为 $u(x,\psi_0,\psi)$，将 $u(x,\psi_0,\psi)$ 代入哈密顿系统后得到的系统，即

$$\begin{cases}\dot{x}=\dfrac{\partial H}{\partial\psi^{\mathrm{T}}}\bigg|_{u=u(x,\psi_0,\psi)}=f(x,u(x,\psi_0,\psi)),\\ \dot{x}_0=f^0(x,u(x,\psi_0,\psi)),\\ \dot{\psi}=-\dfrac{\partial H}{\partial x}\bigg|_{u=u(x,\psi_0,\psi)}\\ \quad=-\left(\dfrac{\partial f}{\partial x}\right)^{\mathrm{T}}\bigg|_{u=u(x,\psi_0,\psi)}\psi-\left(\dfrac{\partial f^0}{\partial x}\right)^{\mathrm{T}}\bigg|_{u=u(x,\psi_0,\psi)}\psi_0,\\ \dot{\psi}_0=0\end{cases}$$

称此为"极大值原理"条目中(1),(2)的极值控制系统. 极值控制系统的以哈密顿系统的端点条件为端点条件的解记为 $x_e(t),\psi_{0e}(t),\psi_e(t)$，且

$$u_e(t)\triangleq u(x_e(t),\psi_{0e}(t),\psi_e(t)),$$

称 $u_e(t)$ 为"极大值原理"条目中(1),(2)的极值控制，$x_e(t)$ 称为极值轨线. 极值控制不一定是最优控制，但是最优控制的"候选者". 当使哈密顿函数 $H(x,\psi_0,\psi,u)$ 关于 u 达到极大的 u 存在、惟一，而且又知最优控制存在时，极值控制就是最优控制.

最优控制综合(optimal control synthesis)　对控制系统构造(寻找)使其对相应性能指标取最小同时依赖时间和状态的控制函数的过程. 当最优控制 $u^*(t)$ 只是时间 t 的函数，称之为程序控制或开环控制. 对于受控制系统

$$\dot{x}=f(t,x,u) \tag{1}$$

和性能指标

$$J[u(\cdot)]=\int_{t_0}^{t_f}f^0(t,x(t),u(t))\mathrm{d}t, \tag{2}$$

还有一种同时以时间 t 和状态 x 为其变元的控制函数 $u(t,x)$. 如果控制函数 $u(t,x)$ 具有如下性质：对任意 $x_0\in\mathrm{R}^n$(或 x_0 属于 R^n 中某区域)，由

$$u^*(t)\triangleq u(t,x^*(t))$$

构成的以时间 t 为其变元的控制函数是(1),(2)的分别以 x_0 和 x_f 为初始状态和终端状态的最优控制，则称 $u(t,x)$ 为(1),(2)的最优控制综合函数，简称最优控制综合函数，这里 $x^*(t)$ 是闭环系统

$$\dot{x}=f(t,x,u(t,x)) \tag{3}$$

满足 $x^*(t_0)=x_0,x^*(t_f^*)=x_f$ 的解，且 $x^*(t)$ 正是(1),(2)对应于 $u^*(t)$ 的轨线即最优轨线. 最优控制综合函数是一种状态反馈形式的最优控制函数. 系统(3)称为(1),(2)的最优闭环系统，简称最优控制系统，最优控制系统总是按最优方式运行.

时间最优控制(time optimal control)　使系统过渡时间(由初态转移到终端状态所需要的时间)取最小的控制. 当"最优控制"条目中的性能指标(2)中标量函数 $f^0(x,u)\equiv 1$ 时，$J[u(\cdot)]=t_f-t_0$ 表示受控系统从 x_0 转移到 x_f 所需时间，这时的最优控制问题乃是在容许控制集合 $\mathscr{U}_{ad}$ 中寻找一个控制，它使系统从 x_0 转移到 x_f 所需时间最小，即时间最优控制问题，通常也称为快速控制问题. 线性定常受控系统的时间最优控制问题的基本结果如下：系统满足端点条件和性能指标分别为

$$\dot{x}=Ax+Bu, \tag{1}$$

$$x(t_0)=x_0,\ x(t_f)=0, \tag{2}$$

$$J[u(\cdot)]=t_f-t_0. \tag{3}$$

(1)式中 A 为 $n\times n$ 常数矩阵，$B=[b_1,b_2,\cdots,b_r]$，b_i 为 n 维常向量，$i=1,2,\cdots,r$，$u=[u_1,u_2,\cdots,u_r]^{\mathrm{T}}$，$r\leqslant n$. 当下列条件：

1. U_r：$|u_i|\leqslant\alpha_i,\alpha_i>0\ (i=1,2,\cdots,r)$.
2. $\mathscr{U}_{ad}$ 非空.
3. $\mathrm{rank}[b_i,Ab_i,\cdots,A^{(n-1)}b_i]=n\ (i=1,2,\cdots,r)$

成立时，系统把 $x_0\in\mathrm{R}^n$ 转移到 $x_f=0$ 的时间最优控制 $u^*(t)=[u_1^*(t),u_2^*(t),\cdots,u_r^*(t)]^{\mathrm{T}}$ 存在、惟一，并且具有如下形式

$$\begin{gathered}u_i^*(t)=-\alpha_i\cdot\mathrm{sgn}(\mu^{\mathrm{T}}\mathrm{e}^{A(t_f^*-t)}b_i)\\ (i=1,2,\cdots,r),\\ -1+\mu^{\mathrm{T}}Bu^*(t_f^*)=0,\end{gathered} \tag{4}$$

其中 μ 为待求的 n 维向量，$\mathrm{sgn}(\cdot)$ 表示$(\cdot)$的符号函数，即

$$\mathrm{sgn}(z)=\begin{cases}1 & (z>0),\\ -1\text{ 或 }+1 & (z=0),\\ -1 & (z<0).\end{cases}$$

亦称条件 3 为最广位置条件，俗称非退化条件(参见"最广位置条件").

最广位置条件(general position condition)　研究线性定常控制系统时间最优控制问题时涉及的条

件. 令 A 和 B 分别表示所论系统的 $n\times n$ 阶系统矩阵和 $n\times r$ 阶控制矩阵, U 表示 r 维欧氏空间 $\mathbb{R}^r$ 中的有界多面体, 对于平行于 U 的每一条棱的向量 ω, 如果向量组 $B\omega, AB\omega, \cdots, A^{(n-1)}B\omega$ 线性无关, 则称线性定常系统满足最广位置条件. 特别地, 记 $B=[b_1,b_2,\cdots,b_r]$, 当 U 为 r 维欧氏空间中对称于原点立方体时, 下列 r 个关系式

$$\operatorname{rank}[b_i, Ab_i, \cdots, A^{(n-1)}b_i] = n \quad (i=1,2,\cdots,r)$$

组成线性定常系统满足的最广位置条件.

bang-bang **控制原理**(bang-bang control principle) 一种快速控制的控制原理. 当非退化条件成立时, 从"时间最优控制"条目中 $u_i^*(t)$ 的表达式(4)看出, 对于任意常向量 μ, 标量函数

$$\varphi_i(t) = \mu^{\mathrm{T}} \mathrm{e}^{A(t_i^*-t)} b_i \quad (i=1,2,\cdots,r),$$

在任意有限时间区间上最多只有有限多个零点, 即时间最优控制函数的每个分量 $u_i^*(t)$ 在任意有限时间区间上除 $\varphi_i(t)$ 的零点外处处有定义, 并且 $u_i^*(t)$ 是分段取常值 α_i 或 $-\alpha_i$ 的函数, $i=1,2,\cdots,r$. $u^*(t)$ 是仅在长方体 U 的顶点处取值的分段常值向量函数. $u_i^*(t)$ 仅在 $\varphi_i(t)$ 的零点处不连续. 时间最优控制的不连续点称为开关点. 使 $\varphi_i(t)$ 为零的时刻称为最优控制的开关时刻. 记 t' 为 $u^*(t)$ 的开关时刻, 且

$$u^*(t'_{-0}) = e_k, \quad u^*(t'_{+0}) = e_l,$$

其中 e_k, e_l 为 U 的不同顶点. 在开关时刻 t', 时间最优控制 $u^*(t)$ 从 U 的一个顶点转移到另一个顶点. 英文表达这种顶点间跳来跳去以 bang-bang 的声音描述, 中文亦以音译. 当 U_r 为 r 维欧氏空间 $\mathbb{R}^r$ 中的凸多面体, 且 $\mathscr{U}_{\mathrm{ad}}$ 非空, 对应的"最广位置条件"成立时, "时间最优控制"条目中系统(1)的把 $\mathbb{R}^n$ 中任一状态 $x_0\neq 0_n$ 控制到 $x_f=0_n$ 的时间最优控制存在、惟一, 且时间最优控制是仅在凸多面体 U_r 的顶点处取值的分段常值向量函数, 同时 bang-bang 控制原理成立.

开关次数定理(theorem on number of switches) 关于时间最优控制开关(切换)次数的定理. 当"时间最优控制"条目中的条件 1,2,3 同时成立, 又 $n\times n$ 矩阵 A 的特征值皆为实数, 这时标量函数

$$\varphi_i(t) = \mu^{\mathrm{T}} \mathrm{e}^{A(t_i^*-t)} b_i \quad (i=1,2,\cdots,r)$$

最多只有 $n-1$ 个零点, 即时间最优控制 $u^*(t)$ 的每个分量 $u_i^*(t)$ 最多只开关 $n-1$ 次. 对于所论系统为标量输入, 即单输入受控系统的情况, 即 $B=b, u$ 为标量的情况, 条件 1,3 分别为

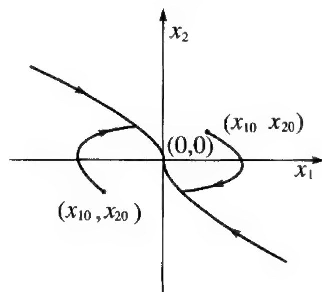
$$|u| \leqslant \alpha, \ \alpha > 0,$$
$$\operatorname{rank}[b, Ab, \cdots, A^{(n-1)}b] = n.$$

如果 A 的特征值皆是实数, 那么对于任意 $x_0(\neq 0)\in\mathbb{R}^n$, 分别取值 α 和 $-\alpha$ 的快速控制函数, 最多经过 $(n-1)$ 开关, 即可把 x_0 最快地控制到 $x_f=0$.

开关曲线(switching curve) 使时间最优控制函数改变符号的曲线. 二阶时间最优控制系统的状态空间是二维欧几里得平面 $\mathbb{R}^2$(或是 $\mathbb{R}^2$ 中包含原点为内点的某区域 Ω^2), 对于控制作用是标量的情况, 如果一条通过坐标原点的曲线 L 把相平面 $\mathbb{R}^2$ 分为两部分 $\mathbb{R}^2_+$ 和 $\mathbb{R}^2_-$(或把 Ω^2 分为两部分 Ω^2_+ 和 Ω^2_-), 即 $\mathbb{R}^2=\mathbb{R}^2_+\cup L\cap\mathbb{R}^2_-$(或 $\Omega^2=\Omega^2_+\cup L\cup\Omega^2_-$), 使得在 $\mathbb{R}^2_+$(或 Ω^2_+)内最优控制取正值, 在 $\mathbb{R}^2_-$(或 Ω^2_-)内最优控制取负值, 在曲线 L 上, 最优控制的取值亦是完全确定的, 并且经过原点的那段 L 必定是时间最优轨线, 则称曲线 L 是时间最优控制系统的开关曲线. 当控制作用是二维向量时, 时间最优控制系统有两条开关曲线. 这样, 二阶线性定常受控系统的时间最优控制的综合在于求得其开关曲线.

双积分环节系统最速控制(time-optimal control for double-integral system) 寻求双积分环节系统的时间最优控制综合函数的问题. 双积分环节系统方程为

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = u. \end{cases}$$



如刚体某个姿态角受控运动的简化动力学模型是双积分环节系统, 其中 u 可看成驱动姿态角运动的外力矩, 视为控制量, 其大小受 $|u|\leqslant M$ (M 为正常数)的约束. 双积分环节的时间最优控制综合函数

$$u^*(x_1,x_2) = \begin{cases} M, & S(x_1,x_2)=0, \ x_2<0, \\ -M\operatorname{sgn}S(x_1,x_2), & S(x_1,x_2)\neq 0, \\ -M, & S(x_1,x_2)=0, \ x_2>0, \end{cases}$$

$$S(x_1,x_2) \triangleq x_1 + \frac{x_2^2}{2M}\operatorname{sgn} x_2$$

表示开关函数, 即 $S(x_1,x_2)=0$ 是双积分环节系统的开关曲线. 双积分环节系统时间最优控制综合函数的相图如图所示. 图中给出了从相平面上任一点 (x_{10},x_{20}) 开始时间最优轨线运动的方式: 最多开关一次即可到达原点. 图中箭头表示运动的方向.

线性二次最优控制(linear quadratic optimal control) 受控系统是线性、性能指标为二次型积分时的最优控制. 简称 LQ 最优控制. LQ 最优控制问题的数学表达如下:

$$\begin{cases} \dot{x} = A(t)x + B(t)u, \\ x(t_0) = x_0 \in \mathbb{R}^n, \ x(t_f) \text{ 自由}, \\ \min\limits_{u\in\mathscr{U}_{\mathrm{ad}}} J[u(\cdot)] = \min\limits_{u\in\mathscr{U}_{\mathrm{ad}}} \dfrac{1}{2}\displaystyle\int_{t_0}^{t_f} [x^{\mathrm{T}}(t)Q(t)x(t) \\ \qquad + u^{\mathrm{T}}(t)R(t)u(t)]\mathrm{d}t, \end{cases}$$

这里 $x\in\mathbb{R}^n, u\in\mathbb{R}^r, r\leqslant n; A(t)=[a_{ij}(t)], B(t)=[b_{ik}(t)], Q(t)=[q_{ij}(t)], R(t)=[r_{lk}(t)]$ 分别是定义在区间 $[t_0,t_f]$ 上的具有相应维数的矩阵值函数，且 $Q(t)$ 为半正定对称矩阵值函数，$R(t)$ 为正定对称矩阵值函数，$\forall t\in[t_0,t_f]$. 通常假定 $a_{ij}(t), b_{ik}(t), q_{ij}(t), r_{lk}(t)$ 都是定义在 $[t_0,t_f]$ 上的分段连续函数. t_0, t_f 可以事先给定，亦可待定. 当 t_0, t_f 为事先给定时，线性二次最优控制问题的解存在、惟一，并且其最优控制综合函数 $u(t,x)$ 具有如下形式：

$$u(t,x)=-R^{-1}(t)B^{\mathrm{T}}(t)P(t)x,$$

上式中 $P(t)$ 是下列黎卡提矩阵微分方程终值问题

$$\begin{cases}-\dot{P}=PA(t)+A^{\mathrm{T}}(t)P+Q(t)\\ \qquad -PB(t)R^{-1}(t)B^{\mathrm{T}}(t)P,\\ P(t_f)=0\ (0\text{ 为 }n\times n\text{ 零矩阵})\end{cases}$$

的半正定对称解矩阵.

线性二次最优调节（linear quadratic optimal regulation） 线性系统在无穷时间区间 $[0,+\infty)$ 上二次型积分性能指标下的最优控制，即

$$\begin{cases}\dot{x}=Ax+Bu,\\ x(t_0)=x_0\in\mathbb{R}^n,\\ \min\limits_{u(\cdot)\in\mathscr{U}_{ad}} J[u(\cdot)]=\min\limits_{u(t)\in\mathscr{U}_{ad}}\dfrac{1}{2}\displaystyle\int_0^{\infty}[x^{\mathrm{T}}(t)Qx(t)\\ \qquad\qquad +u^{\mathrm{T}}(t)Ru(t)]\mathrm{d}t,\end{cases}$$

这里 A,B,Q,R 皆是具有相应维数的常数矩阵；$Q=C^{\mathrm{T}}C\geqslant 0$ 与 $R=R^{\mathrm{T}}>0$ 分别是半正定与正定对称矩阵. 线性二次最优调节问题是指寻找所论线性二次最优控制问题的具有状态反馈形式的最优控制综合函数，并使相应的闭环系统是李亚普诺夫意义下渐近稳定的. 在一定条件下，线性二次最优调节问题的解存在、惟一，并且其最优控制综合函数 $u(x)$ 具有如下形式：$u(x)=-R^{-1}B^{\mathrm{T}}Px$，其中 P 是黎卡提矩阵代数方程

$$PA+A^{\mathrm{T}}P+Q-PBR^{-1}B^{\mathrm{T}}P=0$$

的惟一半正定对称矩阵解，这里 0 是 $n\times n$ 零矩阵.

线性二次最优调节系统（linear quadratic optimal requlation system） 按无穷时间区间上二次型积分性能指标最优运行的线性定常系统. 当 (A,B) 能控、(A,C^{T}) 能观测时，"线性二次最优调节"条目中黎卡提矩阵代数方程存在惟一正定对称解矩阵 P^*，这里 $Q=C^{\mathrm{T}}C$. 在最优控制综合函数

$$u^*(x)=-R^{-1}B^{\mathrm{T}}P^*x\triangleq -K^*x$$

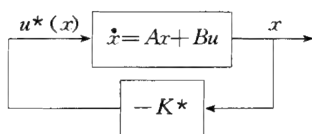
作用下形成的最优闭环系统

$$\dot{x}=[A-BR^{-1}B^{\mathrm{T}}P^*]x,$$

称为线性二次最优调节系统，其框图如图所示. 最优控制综合函数

$$u^*(x)=-R^{-1}B^{\mathrm{T}}P^*x$$

又称为最优调节器. 线性二次最优调节系统具有如下性质：



1. $x=0$ 是它的惟一渐近稳定平衡状态.

2. 二次型标量函数

$$V(x)=\frac{1}{2}x^{\mathrm{T}}P^*x$$

是它的一个李亚普诺夫函数.

3. 它的从任一 $x_0\in\mathbb{R}^n$ 出发，衰减到球心位于平衡点 $x=0$、半径为 $\varepsilon(\varepsilon>0)$ 的球面 $\mathscr{S}_\varepsilon(0,\varepsilon): x^{\mathrm{T}}x=\varepsilon^2$ 的衰减时间 $t_\varepsilon-t_0$ 的估计式为

$$t_\varepsilon-t_0\leqslant-\lambda_+(P^*)\ln\frac{\lambda_-(P^*)}{x_0^{\mathrm{T}}P^*x_0}\varepsilon^2,$$

$\lambda_-(P^*), \lambda_+(P^*)$ 分别是 P^* 的最小、最大特征值.

线性二次最优调节逆问题（inverse problem of linear quadratic optimal regulation） 寻求给定闭环系统是否存在二次型积分指标，使该系统为对应指标下的最优调节系统的问题. 设有线性定常系统 $\dot{x}=Ax+Bu$. 线性二次最优调节逆问题指：对于给定的线性反馈控制规律 $u=-Kx$，它使相应的闭环系统 $\dot{x}=[A-BK]x$ 渐近稳定，寻找 $n\times n$ 半正定对称矩阵 Q 及 $r\times r$ 正定对称矩阵 R，使得 $u=-Kx$ 恰是对应于 Q,R 的二次型积分性能指标

$$J[u(\cdot)]=\frac{1}{2}\int_0^{+\infty}[x^{\mathrm{T}}(t)Qx(t)+u^{\mathrm{T}}(t)Ru(t)]\mathrm{d}t$$

的最优调节器. 假定 $[A,B]$ 能控、$[A,K]$ 能观测，且 $A-BK$ 为稳定矩阵. 如果对于给定的 $r\times r$ 正定对称矩阵 R，四个矩阵 A,B,K 和 R 满足下列矩阵不等式：

$$[I_r-R^{\frac{1}{2}}K(-\mathrm{j}\omega I_n-A)^{-1}BR^{-\frac{1}{2}}]^{\mathrm{T}}$$
$$\cdot[I_r-R^{\frac{1}{2}}K(\mathrm{j}\omega I_n-A)^{-1}BR^{-\frac{1}{2}}]\geqslant I_r,$$

等号仅对有限多个 ω 成立，则存在半正定对称矩阵 Q，使得 $u=-Kx$ 恰是对应于这对 Q,R 构成的二次型积分性能指标的最优调节器. 对于单输入系统，即 $B=b$ 为 n 维向量，u 为标量控制的情况. 这时 R 为正常数，不妨取为 1，上列矩阵不等式成为如下不等式：

$$|1-k^{\mathrm{T}}(\mathrm{j}\omega I_n-A)^{-1}b|\geqslant 1,$$

等号仅对有限多个 ω 成立. 此不等式称为调节器 $u(x)=-k^{\mathrm{T}}x$ 的频域条件，由此可知，对于线性定常能控的单输入系统，仅当使闭环稳定的反馈增益向量 k^{T} 与系统矩阵 A 构成能观测对 $[A,k^{\mathrm{T}}]$，且以 k^{T} 为开环增益的奈奎斯特图满足频域条件时，k^{T} 才是对应于某个二次型积分性能指标的最优增益行向量.

最优调节器的频域条件（frequency-domain condition of optimal regulator） 刻画单输入线性

二次最优调节器的频域特征的条件. 单输入线性定常系统,其二次最优反馈增益(行)向量 $k^{*\mathrm{T}} \triangleq b^{\mathrm{T}}P^*$ 满足不等式

$$|1-k^{*\mathrm{T}}(\mathrm{j}\omega I_n - A)^{-1}b| \geqslant 1$$
$$(-\infty < \omega < +\infty).$$

此不等式称为最优调节器 $u^*(x)=-k^{*\mathrm{T}}x$ 的频域条件. 单输入线性二次最优反馈系统可视为具有单位负反馈的闭环系统,即其开环频率特性是

$$W_0(\mathrm{j}\omega) = k^{*\mathrm{T}}(\mathrm{j}\omega I_n - A)^{-1}b.$$

最优调节器的频域条件表明:对于单输入线性二次最优反馈系统,其开环频率特性 $W_0(\mathrm{j}\omega)$ 的奈奎斯特曲线总是位于复平面上以$(-1,0)$为圆心、单位长度为半径的圆的外部. 这样,一个能控的单输入线性定常系统,对于给定的稳定线性反馈控制规律 $u(x)=-k^{\mathrm{T}}x$,当$[A,k^{\mathrm{T}}]$能观测时,$u(x)=-k^{\mathrm{T}}x$ 构成该系统的二次最优调节器的充分必要条件是频域条件成立.

黎卡提矩阵代数方程的迭代解法(iterative method for Riccati matrixalgebraic equation) 求解黎卡提矩阵代数方程的一种方法. 黎卡提矩阵代数方程

$$PA + A^{\mathrm{T}}P + Q - PBR^{-1}B^{\mathrm{T}}P = 0 \qquad (1)$$

的求解关系到许多控制问题. 由于它是非线性矩阵代数方程,通常不易求得其矩阵解的解析表达式. 这里介绍求解方程(1)的正定对称矩阵解的迭代算法. 假定$[A,B]$能控,$[A,C^{\mathrm{T}}]$能观测,$Q=C^{\mathrm{T}}C$. 迭代算法如下:

1. 取 $r\times n$ 矩阵 K_0,使 $A-BK_0$ 的特征值均具有负实部,以保证李亚普诺夫矩阵代数方程

$$P(A-BK_0) + (A-BK_0)^{\mathrm{T}}P + Q_0 = 0 \quad (2)$$

存在惟一正定对称矩阵解 P_0,其中

$$Q_0 = Q + K_0^{\mathrm{T}}RK_0 \triangleq C_0^{\mathrm{T}}C_0.$$

2. 记 $A_1 = A - BK_1$,$K_1 \triangleq R^{-1}B^{\mathrm{T}}P_0$,$Q_1 = Q + K_1^{\mathrm{T}}RK_1 \triangleq C_1^{\mathrm{T}}C_1$,求李亚普诺夫矩阵方程

$$PA_1 + A_1^{\mathrm{T}}P + Q_1 = 0 \qquad (3)$$

的正定对称阵解 P_1.

3. 对任意正整数 $m(m>1)$,求李亚普诺夫矩阵代数方程

$$P(A-BK_m) + (A-BK_m)^{\mathrm{T}}P + Q_m = 0 \quad (4)$$

的惟一正定对称矩阵解 P_m,这里

$$A_m \triangleq A - BK_m,\ K_m \triangleq R^{-1}B^{\mathrm{T}}P_{m-1},$$
$$Q_m = Q + K_m^{\mathrm{T}}RK_m \triangleq C_m^{\mathrm{T}}C_m.$$

4. 对事先给定的精度要求 $\varepsilon>0$,若 $\|P_m - P_{m-1}\| \leqslant \varepsilon$,则取 P_m 为(1)的正定对称矩阵解逼近解矩阵,否则,以 P_m 代替 P_{m-1}重复迭代算法 3 得 P_{m+1}.

黎卡提矩阵代数方程的哈密顿解法(Hamilton method for Riccati matrix algebraic equation) 求解黎卡提矩阵代数方程的一种方法. 仍考虑"黎卡提矩阵代数方程的迭代解法"条目中的黎卡提矩阵代数方程(1). 首先,组成哈密顿矩阵

$$H = \begin{pmatrix} A & -BR^{-1}B^{\mathrm{T}} \\ -Q & -A^{\mathrm{T}} \end{pmatrix},$$

能够证明矩阵 H 没有纯虚数特征值,且若 λ 是它的特征值,则$-\lambda$ 也是它的特征值,并且两者的重数相同. 因而,按照舒尔定理,存在 $2n\times 2n$ 的可逆矩阵 T,使得

$$S = T^{-1}HT = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix},$$

其中 S_{11}的特征值均具负实部,而 S_{22}的特征值均具正实部. 若将上述变换矩阵 T 分块为

$$\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix},$$

则能够证明 T_{11}是可逆的,且方程(1)的正定对称矩阵解 $P=T_{21}T_{11}^{-1}$.

黎卡提矩阵代数方程的符号函数解法(sign function method for Riccati matrix algebraic equation) 求解黎卡提矩阵代数方程的一种方法. 仍考虑"黎卡提矩阵代数方程的迭代解法"条目中黎卡提矩阵代数方程(1),并组成哈密顿解法中的哈密顿矩阵 H. 矩阵 H 没有纯虚数特征值. 若它的若当形为对角形,则存在可逆矩阵 T,使 $H=T\Lambda T^{-1}$,其中 $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_{2n})$. 矩阵 H 的符号函数是

$$S_H = \mathrm{sgn}(H)$$
$$= T\mathrm{diag}(\mathrm{sgn}(\lambda_1),\cdots,\mathrm{sgn}(\lambda_{2n}))T^{-1},$$

其中复数 λ_i 的符号函数是

$$\mathrm{sgn}(\lambda_i) = \begin{cases} +1, & \text{当 } \lambda_i \text{ 的实部大于 } 0, \\ -1, & \text{当 } \lambda_i \text{ 的实部小于 } 0. \end{cases}$$

计算 S_H 的一种方法是

$$H_0 = H,\ H_{i+1} = \alpha H_i + \beta H_i^{-1} \quad (i=0,1,2,\cdots).$$

能够证明 $i\to\infty$时,$H_i \to S_H$,其中 α,β 的一种取法是 $\alpha=\beta=1/2$. 在得到 H 的符号函数 S_H 后,可以确定它的符号正函数

$$W = \mathrm{sgn}^+(H) = \frac{1}{2}(\mathrm{sgn}(H) + I).$$

将矩阵 W 分块为

$$\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix},$$

则能够证明

$$\begin{pmatrix} W_{12} \\ W_{22} \end{pmatrix}$$

是满秩的，且黎卡提矩阵代数方程(1)的正定对称矩阵解 P 为

$$P=-(W_{12}^2+W_{22}^2)^{-1}(W_{12}W_{11}+W_{22}W_{21}).$$

具有指定衰减度的线性二次最优调节(linear quadratic optimal regulation with assigned decresing rate)　在二次性能指标中考虑了对衰减度要求时的最优调节. 讨论最优调节问题：

$$\begin{cases}\dot{x}=Ax+Bu,\\ x(t_0)=x_0\in \mathrm{R}^n,\\ \min\limits_{u(t)\in\mathscr{U}_{\mathrm{ad}}} J[u(\cdot)]=\min\limits_{u(t)\in\mathscr{U}_{\mathrm{ad}}}\dfrac{1}{2}\displaystyle\int_{t_0}^{\infty}\mathrm{e}^{2\alpha t}[x^{\mathrm{T}}(t)Qx(t)\\ \qquad\qquad +u^{\mathrm{T}}(t)Ru(t)]\mathrm{d}t,\end{cases}$$

这里 α 是事先指定的正常数，其余记号见“线性二次最优调节”. 具有指定衰减度 α 的线性二次最优调节问题是指寻找上列最优控制问题的最优控制综合函数，它使得相应的闭环系统成为具有指定衰减度 α 的渐近稳定系统. 记 $Q=C^{\mathrm{T}}C$，当 (A,B) 能控，(A,C^{T}) 能观测时，具有指定衰减度 α 的线性二次最优调节问题的解存在、惟一，且其最优控制综合函数 $u_\alpha^*(x)$ 具有如下形式：$u_\alpha^*(x)=-R^{-1}B^{\mathrm{T}}P_\alpha^*x$，其中 P_α^* 是下列黎卡提矩阵代数方程

$$P(A-\alpha I_n)+(A-\alpha I_n)^{\mathrm{T}}P$$
$$+Q-PBR^{-1}B^{\mathrm{T}}P=0$$

的惟一正定对称矩阵解. 在 $u_\alpha^*(x)$ 作用下形成的闭环系统

$$\dot{x}=[A-BR^{-1}B^{\mathrm{T}}P_\alpha^*]x$$

称为具有指定衰减度 α 的最优调节系统. 最优控制综合函数 $u_\alpha^*(x)=-R^{-1}B^{\mathrm{T}}P_\alpha^*x$ 称为具有指定衰减度的最优调节器. 可以证明下列事实成立：

1. 具有指定衰减度 α 的最优调节系统以 $x=0$ 为其惟一的渐近稳定平衡状态.

2. 系统 $\dot{x}=[A-\alpha I_n-BR^{-1}B^{\mathrm{T}}P_\alpha^*]x$ 以 $x=0$ 为其惟一的渐近稳定平衡状态.

3. 对任意 $x_0\in\mathrm{R}^n$，具有指定衰减度 α 的最优调节系统与事实 2 中系统的以 x_0 为初始状态的解分别记为 $x_\alpha^*(t)$ 与 $x_\alpha(t)$，这时 $x_\alpha^*(t)$ 与 $x_\alpha(t)$ 之间成立下列关系式 $x_\alpha^*(t)=\mathrm{e}^{-\alpha t}x_\alpha(t)$.

跳跃条件(jump condition)　状态受限最优控制系统的最优轨线段的连结点所满足的一个条件. 在许多最优控制系统中，虽然控制参数的大小范围已经受到限制，但这样的控制量有时可将系统状态控制到从安全性和可靠性观点来看是完全不允许的地方，因而对系统状态也要给出一定的限制. 通常将状态限制在状态空间内一个具有光滑边界的有界闭集合 G 上，这时控制系统的最优轨线由两部分组成. 一部分位于 G 的内部，一部分位于 G 的边界上. 要完整地将整个轨线确定出来，除弄清楚每部分轨线段外，还要搞清楚两部分轨线连结点的情况. 跳跃条件正是这两部分轨线连结点所满足的条件.

追踪时间(pursuit time)　追踪控制系统的最小相遇时间. 在 n 维欧氏空间内有两个受控系统 Ⅰ 和 Ⅱ 在运动，系统 Ⅰ 的容许控制集合与状态向量函数分别记为 $\mathscr{U}_{\mathrm{ad}}^1$ 与 $x(t)$；系统 Ⅱ 的容许控制集合与状态向量函数分别记为 $\mathscr{U}_{\mathrm{ad}}^2$ 与 $y(t)$. 对任给的 $v(t)\in\mathscr{U}_{\mathrm{ad}}^2$，若存在系统 Ⅰ 的控制 $u(t)\in\mathscr{U}_{\mathrm{ad}}^1$，使系统 Ⅰ 和系统 Ⅱ 在 t_1 时刻相遇，即相应的状态 $x(t)$ 与 $y(t)$ 满足 $x(t_1)=y(t_1)$，则称 t_1 为一个相遇时间. 可能 $\mathscr{U}_{\mathrm{ad}}^1$ 中有另外的控制函数，它使两系统相遇. 这些相遇时间的最小者称为追踪时间.

时滞控制系统(time delay control system)　基于信号的延迟性而形成的受控系统. 控制系统工作的基础是各种信号的传输，由于在实际系统中，信号的传输和再现都不是即时的，而往往要有一段时间滞后，这就是系统的滞后效应. 在系统建模时基于这种考虑而形成的控制系统即是时滞控制系统. 例如，由状态方程 $\dot{x}=f(t,x(t),x(t-\theta),u(t)),t\in[t_0,t_1]$ 描述的受控系统即是一个时滞控制系统，式中 $x(t)\in\mathrm{R}^n$ 是状态向量，$u(t)\in\mathrm{R}^r$ 是控制向量，常数 θ 是时间滞后.

动态规划(dynamic programming)　求解多步时间决策过程的方法. 贝尔曼(Bellman, R.)于 20 世纪 50 年代中期提出. 用动态规划方法处理决策过程必须满足两个基本条件：过程的马尔可夫性或无后效性和指标函数的可分性. 动态规划方法的基本思想是“以空间换时间”. 其处理问题的框架程序是利用贝尔曼提出的“最优性原理”，将问题剖分成若干子问题，然后逐次迭代求出整个问题的最优解. 因而，用动态规划方法处理问题能够节省大量的计算时间. 但是，此方法在迭代过程中需要较大的计算机内存空间，因而此方法处理高维问题时会遇到“维数障碍”. 这里的“维数”不是指决策过程状态变量的向量维数，而是指状态变量的个数. 用动态规划方法处理实际问题时，如何确定问题的状态变量是应用此方法的关键技巧之一. 现已有了某些克服所谓“维数障碍”的降维法，如拉格朗日降维法. 在最优控制理论中，动态规划方法是求解最优控制问题的有效手段之一. 动态规划方法还有很多实际应用，如它能有效求解网络与图的最短路径问题、存储问题和数学规划问题等.

连续决策过程的构成要素(elements of continuous making-decision process)　构成动态规划模型的确定性集中参数连续时间决策过程的必要内容. 由下列五个要素构成：

1. 时间段要素. 由时间变量集合组成，时间变量集合是非负实数轴上的区间 T，它表示决策的长度，

称为历程.历程可以有限,亦可无穷.

2.状态要素.由有穷多个有序数组组成,属于有穷维欧几里得空间 $\mathbb{R}^n$(或 $\mathbb{R}^n$ 中某区域 χ_n),它表示决策过程的状态演变特性.

3.决策、策略规则与策略.无反馈决策随时间演化.用 T 到策略取值范围 $U_r\subset\mathbb{R}^r$ 的函数 $u(t)$表示.决策规则用在给定时刻 $t\in T$ 定义在某个 $\chi_n\subset\mathbb{R}^n$ 上的函数 $u_t(x)$表示.反馈策略 P 用定义在 $T\times\chi_n$ 上的函数 $u(t,x)$表示.用 $\mathscr{U}\triangleq\{u(t,x)\mid,t\in T,x\in\chi_n\}$ 表示容许决策集.

4.状态转移规律.指决策过程的演变特征.由向量常微分方程 $\dot{x}=f(t,x,u)$表示.假定对给定的初始状态 $x(t_0)=x_0$ 和定义在$[t_0,t_f]$$(t_0<t_f)$上的决策 $u(t)$(或 $u(t,x)$),微分方程的解 $x(t)$在$[t_0,t_f]$上存在、惟一.这时对$[t_0,t_f]$上任意两个时刻 $t_1,t_2,t_1<t_2$,相应的状态 $x(t_1),x(t_2)$具有确定的转移关系

$$x(t_2)=\varphi(t_2,t_1,x(t_1),u_{[t_0,t_2]}),$$

其中变换 φ 满足

$$\begin{aligned}&\varphi(t_3,t_1,x(t_1),u_{[t_1,t_3]})\\&=\varphi(t_3,t_2,\varphi(t_2,t_1,x(t_1),u_{[t_1,t_2]}),u_{[t_2,t_3]}),\end{aligned}$$

上式中 $u_{[t_1,t_2]},u_{[t_2,t_3]}$分别表示 $u(t)$(或 $u(t,x(t))$)在区间$[t_1,t_2]$,$[t_2,t_3]$上截取部分.

5.性能指标.是评价决策过程"优劣"的数学描述,通常用下列泛函表示

$$J[u(\cdot)]=\int_{t_0}^{t_f}f^0(t,x(t),u(t))\mathrm{d}t.$$

基于上述五个构成要素,可得到带有固定端点的确定性集中参数连续时间决策过程的动态规划模型如下:

$$\begin{cases}\dot{x}=f(t,x,u),\ x(t_0)=x_0,\ x(t_f)=x_f, & (1)\\ \min\limits_{u(\cdot)\in\mathscr{U}} J[u(\cdot)]=\min\limits_{u(\cdot)\in\mathscr{U}}\displaystyle\int_{t_0}^{t_f}f^0(t,x(t),u(t))\mathrm{d}t,\end{cases}$$

其中 t_f,x_f 固定.当 t_f,x_f 自由或满足其他约束条件时,有带相应端点条件的动态规划模型.

最优性原理(optimality principle) 动态规划理论中的一个基本原理.它由贝尔曼(Bellman,R.)提出,表述为:一最优性策略必具有如下性质:不论初始状态和初始决策如何,对于先前决策造成的状态而言,余下的那些决策必构成一最优策略.就连续决策过程而言,最优性原理的简单叙述是:对于最优决策过程,它的任何最后一段都是最优的.最优性原理是最优决策过程应满足的必要条件.最优性原理是否成立,必须对各个所考虑的动态规划模型分别给予证明.例如,对由"连续决策过程的构成要素"条目中式(1)表示的动态规划模型,能够证明最优性原理成立.

贝尔曼方程(Bellman equation) 带端点条件的确定性集中参数连续时间决策过程的动态规划基本方程."连续决策过程的构成要素"条目(1)中的最优策略、最优轨线和最优性能指标值分别记为 $u^*(t)$,$x^*(t)$和 $J^*=J[u^*(\cdot)]$,J^*是初值(t_0,x_0)的函数,即 $J^*=J^*(t_0,x_0)$.利用最优性原理,能够得到如下结论:如果 $J^*(t,x)$是其变元的连续可微函数,则它满足

$$\begin{cases}\dfrac{\partial J^*(t,x(t))}{\partial t}+\min\limits_{u\in V_r}\{\dfrac{\partial J^*(t,x^*(t))}{\partial x}\\ \quad\cdot f(t,x^*(t),u)+f^0(t,x^*(t),u)\}\\ \quad=\dfrac{\partial J^*(t,x(t))}{\partial t}+\dfrac{\partial J^*(t,x^*(t))}{\partial x}\\ \quad\cdot f(t,x^*(t),u^*(t))\\ \quad+f^0(t,x^*(t),u^*(t))=0,\\ J^*(t_f,x_f)=0.\end{cases}$$

关系式

$$\begin{cases}\dfrac{\partial J(t,x)}{\partial t}+\min\limits_{u\in V_r}\{\dfrac{\partial J(t,x)}{\partial x}f(t,x,u)\\ \quad+f^0(t,x,u)\}=0, & (1)\\ J(t_f,x_f)=0\end{cases}$$

称为带端点条件的贝尔曼方程.

贝尔曼方程的解(solution to Bellman equation) 满足贝尔曼方程的一组函数.假定"连续决策过程的构成要素"条目(1)中的 $f^0(t,x,u)$为定义在$[t_0,t_f]\times\mathbb{R}^n\times U_r$ 上的恒正函数,且 $f^0(t,0,0)=0$.称贝尔曼方程(1)可解,是指下列事实成立:

1.存在连续可微函数 $u\left(t,x,\dfrac{\partial J}{\partial x}\right)$,它使表达式

$$\frac{\partial J}{\partial x}f(t,x,u)+f^0(t,x,u)$$

关于 u 达到极小.

2.下列非线性偏微分方程的终值问题

$$\begin{cases}\dfrac{\partial J}{\partial t}+\dfrac{\partial J}{\partial x}f\left(t,x,u\left(t,x,\dfrac{\partial J}{\partial x}\right)\right)\\ \quad+f^0\left(t,x,u\left(t,x,\dfrac{\partial J}{\partial x}\right)\right)=0,\\ J(t_f,x_f)=0\end{cases}$$

存在解 $J(t,x)$,$J(t_f,x_f)=0$.

称贝尔曼方程(1)具有正解,是指存在一对函数

$$J^*(t,x),\ u^*(t,x)\triangleq u\left(t,x,\frac{\partial J^*(t,x)}{\partial x}\right),$$

$$J^*(t,x)>0,\ \forall t\neq t_f,\ x\neq x_f,\ J^*(t_f,x_f)=0,$$

它们满足

$$\begin{aligned}&\frac{\partial J^*(t,x)}{\partial t}+\frac{\partial J^*(t,x)}{\partial x}f(t,x,u)+f^0(t,x,u)\\ \geqslant&\frac{\partial J^*(t,x)}{\partial t}+\frac{\partial J^*(t,x)}{\partial x}f(t,x,u^*(t,x))\\ &+f^0(t,x,u^*(t,x))=0,\end{aligned}$$

$\forall u\in V_r$,x 属于 $\mathbb{R}^n$ 中 x_f 的某个邻域内.$t\in[t_0,t_f]$,t_0

$<t_1$. 能够证明，当贝尔曼方程(1)具有正解 $J^*(t,x)$，$u^*(t,x)$ 时，$u^*(t,x)$ 组成"连续决策过程的构成要素"条目(1)的最优反馈策略，$J^*(t,x)$ 是相应的以 (t,x) 为初值的最优性能指标值. 一般情况，贝尔曼方程仅存在粘性解(一类广义解).

随机最优控制的极大值原理(maximum principle of stochastic optimal control) 随机控制系统的最优控制应满足的必要条件. 典型随机控制系统的一般描述为

$$\begin{cases} dx(t) = g(x(t),v(t))dt \\ \qquad + \sigma(x(t),v(t))dw(t), \\ x(0) = x_0, \end{cases} \quad (1)$$

其中 w 是布朗运动或维纳过程，v 是 $w(t)$ 适应的控制过程，取值于给定的约束集 U，直观地说，就是 $v(t)$ 可以由 $w(s)$ $(0<s<t)$ 的历史来确定，$x(t)$ 是(1)的解，称为状态轨线，它也是 $w(t)$ 适应的随机过程. 最优控制问题就是求 $v=u(\cdot)$，使泛函

$$J(v(\cdot)) = E\left\{\int_0^T l(x(t),v(t),t) + h(x(T))\right\} \quad (2)$$

取极小值. 这里 E 表示数学期望. 当 $\sigma=0$ 时，(1)，(2)变为确定性系统最优控制问题.

一般随机最优控制的极大值原理叙述如下：在一定的光滑性条件下，如果 $(y(\cdot),u(\cdot))$ 是最优控制问题(1)和(2)的一个解，那么必有

$$(p(\cdot),K(\cdot)) \in L^2_{\mathscr{F}}(0,T;\mathbb{R}^n) \times (L^2_{\mathscr{F}}(0,T;\mathbb{R}^n))^d,$$
$$(P(\cdot),Q(\cdot)) \in L^2_{\mathscr{F}}(0,T;\mathbb{R}^{n\times n}) \times (L^2_{\mathscr{F}}(0,T;\mathbb{R}^{n\times n}))^d,$$

它们分别是下列倒向随机微分方程(3)和(4)的解，并且使得变分不等式(5)成立，此处 $L^2_{\mathscr{F}}$ 是平方可积适应过程所组成的空间；

$$\begin{cases} -dp(t) = [g_x(y(t),u(t))p(t) \\ \qquad + K(t)\sigma_x(y(t),u(t)) \\ \qquad + l_x(y(t),u(t))]dt \\ \qquad - K(t)dw(t), \\ p(T) = h_x(y(T)); \end{cases} \quad (3)$$

$$\begin{cases} -dP(t) = [g_x(y(t),u(t)P(t) \\ \qquad + P(t)g_x(y(t),u(t)) \\ \qquad + P(t)\sigma_x(y(t),u(t)) \\ \qquad + Q(t)\sigma_x(y(t),u(t)) \\ \qquad + H_{xx}(y(t),u(t),p(t), \\ \qquad K(t))]dt - Q(t)dw(t), \\ P(T) = h_{xx}(y(T)); \end{cases} \quad (4)$$

$$\begin{aligned} &H(y(T),u,p(T)K(T), \\ &\quad -P(T)\sigma(y(T),u(T))) \\ &\quad + \frac{1}{2}\mathrm{tr}(\sigma\sigma^*(y(T),u)P(T)) \\ &\geqslant H(y(T),u(T),p(T),K(T) \\ &\quad -P(T)\sigma(y(T),u(T))) \\ &\quad + \frac{1}{2}\mathrm{tr}(\sigma\sigma^*(y(T),u(T))P(T)) \end{aligned} \quad (5)$$
$$(\forall v \in U, \text{a.e.}, \text{a.s.}),$$

其中 p 和 P 分别是一阶和二阶共轭过程，$H(x,v,p,K) = l(x,v) + p \cdot g(x,v) + K \cdot \sigma(x,v)$.

在特殊情况 $\sigma=\sigma(z)$ 下，上述一般极大值原理退化为贝斯缪的最大值原理：

$$H(x_t,p_t,q_t,v) \geqslant H(x_t,p_t,q_t,u_t) \ (\forall v \in U).$$

倒向随机微分方程(backward stochastic differential equation) 一种特殊的随机微分方程. 它的一般形式如下：

$$\begin{cases} -dy(t) = f(y(t),z(t),t)dt - z(t)dw(t), \\ y(T) = \xi, \ 0 \leqslant t \leqslant T, \end{cases} \quad (1)$$

其中 T 是将来的一个给定时刻，ξ 是 T 时刻预定要达到的目标(终端条件)，$w(t)$ 是干扰源(标准的布朗运动). 倒向随机微分方程解的存在惟一性定理是：设 ξ 平方可积并关于 $\mathscr{F}^W_T$ 可测，对于固定的 (y,z)，$f(y,z,\cdot)$ 是 $\mathscr{F}^W_T$ 适应的平方可积过程且关于 (y,z) 满足一致李普希茨条件，则存在惟一的一对平方可积的 $\mathscr{F}^W_T$ 适应过程 $(y(t),z(t))$ 满足倒向随机微分方程(1).

倒向随机微分方程最初的研究起源于随机最优控制的极大值原理中的共轭方程. 倒向随机微分方程与(经典的)正向随机微分方程有很大的不同. 从应用意义上讲，正向随机微分方程告诉人们给定今天的初始状态情况下，明天可能的状态是什么. 而倒向随机微分方程的解 $(y(t),z(t))$ 的含义是在有风险的情况下要想在明天达到某一预定状态，$y(t)$ 告诉人们今天需要具备什么条件，$z(t)$ 则告诉人们如何去做；可以用倒向随机微分方程来描述不确定经济环境下的消费偏好(即效用函数理论)；金融市场的许多重要的衍生证券(如期权、期货等)的理论价格也可以用倒向随机微分方程解出；通过倒向随机微分方程还可获得非线性费曼-卡斯公式，从而可以用来处理拟线性偏微分方程组.

随机线性系统的二次最优控制(quadratic optimal control for stochastic linear system) 性能指标为二次型时的线性随机控制系统的最优控制问题. 为叙述方便起见，这里仅介绍一个简单特例. 设 w_t 为给定概率空间 $(\Omega,\mathscr{F},P,\mathscr{F}_t)$ 上的 m 维布朗运动，$\mathscr{F}_t = \sigma\{w_s: 0 \leqslant s \leqslant t\}$. 考虑线性随机微分方程：

$$\begin{cases} dx = (Ax + Cu)dt + (Bx + Du)dw_t, \\ x(0) = x_0, \end{cases} \quad (1)$$

并极小化下面的二次指标：

$$J(u) = E\left\{\int_0^T |M_t x_t|^2 dt + \int_0^T \langle N_t u_t, u_t \rangle dt + |M_t x_t|^2\right\}, \quad (2)$$

其中，以上所有系数均假设为确定性的有界函数，N_t 正定. 那么随机线性系统二次最优问题（简称 LQ 问题）(1)，(2)的最优解为状态反馈：

$$u_t = -(N + D^{*}PD)^{-1}D^{*}PBx_t, \qquad (3)$$

其中 P 是如下黎卡提矩阵方程终值问题的解：

$$\begin{cases} \dot{P} + PA + A^{T}P + B^{T}PB - (B^{T}PD + PC)(N \\ \quad + D^{T}PD)^{-1}(D^{T}PB + C^{T}P) + M^{T}M = 0, \\ P_T = M_T^{T}M_T. \end{cases}$$

继伯斯姆特（Bismut）的工作后，不少学者对随机 LQ 问题作了大量工作，并且仍在继续之中. 后来人们注意到，二次指标(2)中的 N 可以为负. 从而，N 的非负性假设不再是随机 LQ 问题(1)，(2)有解的必要条件. 这一点与确定性情形下的 LQ 问题存在着本质的差别.

鲁 棒 控 制

鲁棒控制（robust control） 通过控制手段来实现系统的鲁棒性. 鲁棒性是指控制系统性能对于内部、外部变化的不敏感性. 它反映系统运行状态对各种干扰的抵御能力.

一般说来，控制系统中的不确定性可大致分为两类：参数摄动和未建模动态. 前者又称为结构型不确定性（摄动），而后者又称为非结构型不确定性（摄动）. 目前，基于 H_∞ 算子理论的鲁棒设计方法能够有效地处理非结构型摄动，而参数空间方法在处理结构型摄动时比较方便. 结构奇异值方法可以处理非常一般的摄动模式，但其算法的有效性问题还没有彻底解决. H_∞ 控制理论是由加拿大学者詹姆斯（Zames，G.）于 20 世纪 60 年代初提出的，主要目的是在保证闭环系统稳定的前提下，通过反馈作用对外界干扰进行有效的抑制. 后由加拿大的弗朗希斯（Francis，B. A.）、美国的多伊尔（Doyle，J. C.）、英国的哥拉沃（Glover，K.）等人进一步发展，形成了目前基于状态空间的解法，即黎卡提矩阵方程（不等式）解法. 参数空间方法的发展可以追溯到 20 世纪 60—70 年代美国西尔亚克（Siljak，D. D）和德国阿克曼（Ackermann，J.）的工作. 苏联学者哈里托诺夫（Харитонов，В. Л.）在 20 世纪 70 年代末期的创造性工作极大地推进了参数空间方法的发展，他证明了区间多项式的赫尔维茨稳定性可由其四个特殊顶点多项式的赫尔维茨稳定性所保证（无论多项式的阶次有多高），从而大大降低了计算工作量. 随后，各国学者在稳定区域、参数相依性和性能指标等几方面将这一结论进一步推广，建立了菱形（单形）族定理、盒子定理、棱边定理、边界定理、原像定理、根分布定理、凸方向定理、鲁棒小增益定理、鲁棒严格正实定理、鲁棒波波夫定理等. 鲁棒控制（robust control）一词是由加拿大学者戴维森（Davison，E. J.）在 20 世纪 70 年代最先使用的.

μ 理论（μ theory） 为解决控制系统中结构型不确定性所带来的鲁棒性问题而建立的一种理论. 其主要贡献在于引入了结构奇异值（参见“结构奇异值”）. 考虑一个具有多点独立的有界范数摄动的线性系统，通过对系统的输入、输出、传递函数、参数变化、摄动等所有线性关联重构，隔离所有摄动，可将系统化为如图所示的块对角摄动系统. 其中 M 表示广义标称对象，块对角阵 Δ 表示参数或动态不确定性. 若以 $\mu(M)$ 表示结构奇异值，则图示系统对摄动稳定的充分必要条件为

$$\delta \cdot \mu(M) < 1,$$

其中 $\delta = \bar{\sigma}(\Delta)$. μ 理论的特点是：

1. 克服了 H_∞ 理论描述系统结构型不确定性时所带来的保守性.

2. 将结构不确定性下的性能问题转化为鲁棒稳定问题，将鲁棒稳定和鲁棒性能统一到了标准的 μ 问题之下.

关于 μ 的计算十分困难，通常用一些上界、下界的计算来近似逼近它.

闭线性算子（closed linear operator） 一种较特殊的无界算子. 它是惟一可被定量化描述的无界算子. 闭线性算子原是泛函分析中的概念，后被引入鲁棒控制中讨论系统的不稳定摄动问题. 其原始定义为：设 T 为一个 $X \to Y$ 的线性算子，X，Y 分别为两个巴拿赫空间，设 $u_n \in \mathrm{Dom}(T)$ 为 X 中一串元素，所谓$\{u_n\}$是 T 收敛的是指存在 u，使

$$\lim_{n \to +\infty} u_n = u \in X,$$

且存在 v，使

$$\lim_{n \to +\infty} Tu_n = v \in Y,$$

记为 $u_n \underset{T}{\to} v$. 称算子 T 为闭线性算子. 假若 $u_n \underset{T}{\to} u$ 时，有

$$u \in \mathrm{Dom}(T) \text{ 且 } \lim_{n \to +\infty} Tu_n = Tu = v.$$

上面定义的闭线性算子有一个重要性质，即 T 的图象 $G(T) = \{(u, Tu) \mid u \in \mathrm{Dom}(T)\}$ 为乘积空间 $X \times Y$ 的一个闭线性子空间.

后经研究发现，控制系统中一个对象的传递函数 $P(s)$（$n \times m$ 阶实有理矩阵），若仅在有限功率谱输入与输出情况下考虑，实际上等于引入了一个从输入空间 H_2^m 到输出空间 H_2^n 的闭线性算子，这一结论为在鲁棒控制中引入隔扑概念讨论系统的不稳定摄动打下了基础.

算子之间的隔扑（Gap between operators） 泛

函分析中讨论两个闭线性算子之间差别大小的一种概念. 它被引入鲁棒控制理论中,可讨论系统的不稳定摄动的大小. 泛函中的原始定义中首先引入了两个闭线性子空间中的隔扑概念. 设 X 是一个巴拿赫空间,φ 与 ψ 分别为 X 的两个闭线性子空间,令

$$S_\varphi \triangleq \{x \in \varphi \mid \| x \| = 1\},$$

则从 φ 到 ψ 的有向隔扑定义为

$$\vec{\delta}(\varphi,\psi) \triangleq \sup_{x\in S_\varphi} \inf_{y\in\psi} \| x - y \|,$$

而 φ 与 ψ 之间的隔扑被定义为

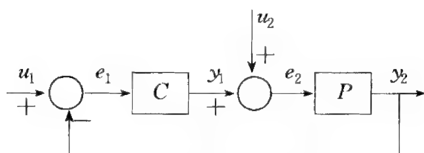
$$\delta(\varphi,\psi) \triangleq \max\{\vec{\delta}(\varphi,\psi), \vec{\delta}(\psi,\varphi)\}.$$

在两个闭线性子空间隔扑定义的基础上,可定义上边两闭线性算子之间的隔扑. 设 T,S 分别为两个由 $X \to Y$ 的闭线性算子,X,Y 分别为两个巴拿赫空间,$G(T),G(S)$ 代表算子 T 与 S 的图象,则两个算子 T,S 之间的隔扑由下式定义

$$\delta(T,S) = \delta(G(T),G(S)).$$

因为在有限功率谱输入和输出的情况下,任一控制系统对象的传递函数 $P(s)$ 可引入一个输入空间到输出空间的一个闭线性算子. 设 $P(s)$ 与 $P_0(s)$ 分别为含摄动对象与名义对象,则这两个对象之间的差别的大小即可用 $P_0(s)$ 与 $P(s)$ 之间的隔扑来度量. 其优点在于即使 $P(s)$ 与 $P_0(s)$ 之间的摄动是不稳定的,也可用隔扑给出一种定量描述,而 H_∞ 理论则无法讨论这一情况.

隔扑理论(Gap theory) 利用闭算子之间的隔扑概念来描述系统不确定性而发展起来的一种有关鲁棒控制的理论. 其主要结果如下. 考虑下图所示的



线性反馈系统. 设 $[P_0,C_0]$ 为图示系统的一个稳定的名义闭环对(亦即名义对象 P_0 和名义控制器 C_0 可构成一个稳定的闭环系统),定义

$$b_{P_0,C_0} \triangleq \left\| \begin{pmatrix} I \\ P_0 \end{pmatrix} (I + C_0P_0)^{-1}[I,C_0] \right\|_\infty^{-1},$$

则可得如下定理:设 $P_0 \in \mathrm{R}^{n\times m}(s)$,$C_0 \in \mathrm{R}^{m\times n}(s)$,则下述命题是等价的:

1. $[P_0,C_0]$ 是稳定的闭环对,且 $b \leqslant b_{P_0,C_0}$.

2. 对所有满足 $\delta(P,P_0) < b$ 的 P,$[P,C_0]$ 是稳定的闭环对.

3. 对所有满足 $\delta(C,C_0) < b$ 的 C,$[P_0,C]$ 是稳定的闭环对.

4. 对所有满足 $\delta(P,P_0) + \delta(C,C_0) < b$ 的 P 和 C,$[P,C]$ 是稳定的闭环对.

这一结果有三个特点:

1. 由此定理可得到一种求最优鲁棒控制器的方法,即设计一个可稳定的 P_0,且使 b_{P_0,C_0} 达到最大的 C_0,从而将隔扑理论化为标准的 H_∞ 问题.

2. 该结果与 H_∞ 理论中用正规互质因子分解描述系统摄动的方法是相通的.

3. 可以推广到分布参数系统.

有限检验(finite verification) 亦称顶点检验. 指系统族的某一性质(如稳定性、动态性能等)可由系统族中的有限个系统的同一性质所保证. 有限检验具有有限检验结论的性质,因而可以在很大程度上减少计算量. 对于区间系统族,有限检验结论也称为哈里托诺夫型结论. 若所需检验顶点系统数目与系统族的阶次无关,则称为强哈里托诺夫型结论;若所需检验的顶点系统数目是系统族阶次的多项式函数,则称该结论为几乎强哈里托诺夫型结论;若所需检验的顶点系统数目是系统族阶次的指数函数,则称该结论为弱哈里托诺夫型结论. 有限检验与哈里托诺夫型结论主要有哈里托诺夫定理、菱形族定理以及关于哈里托诺夫区域和凸方向的研究结果.

顶点检验(vertex verification) 即"有限检验".

哈里托诺夫型结论(Kharitonov like result) 与有限检验有关的结果(参见"有限检验").

哈里托诺夫定理(Kharitonov theorem) 区间多项式族赫尔维茨稳定性(即多项式的根全在左半开复平面)的有限检验定理. 对于 n 次实系数区间多项式族

$$P = \left\{ p(s) = \sum_{i=0}^{n} a_i s^i \mid a_i \in [a_i^-, a_i^+], \right.$$
$$\left. i = 0,1,\cdots,n \right\},$$

全族的稳定性可由下列四个顶点多项式的稳定性所保证

$$p_1(s) = a_0^- + a_1^- s + a_2^+ s^2 + a_3^+ s^3 + \cdots,$$
$$p_2(s) = a_0^- + a_1^+ s + a_2^+ s^2 + a_3^- s^3 + \cdots,$$
$$p_3(s) = a_0^+ + a_1^+ s + a_2^- s^2 + a_3^- s^3 + \cdots,$$
$$p_4(s) = a_0^+ + a_1^- s + a_2^- s^2 + a_3^+ s^3 + \cdots,$$

上述四个多项式通常称为哈里托诺夫多项式. 对于 n 次复系数区间多项式族,全族的稳定性可由特定的八个顶点多项式的稳定性所保证. 这一定理有两个重要特点:

1. 所需检验的顶点数目与多项式族的阶次无关.

2. 只适用于连续时间意义下系统的稳定性判定.

此定理是俄国数学家哈里托诺夫(Харитонов, В. Л.)于 1978 年给出的.

菱形族定理(diamond theorem) 菱形多项式

族赫尔维茨稳定性的有限检验定理. 对于 n 次实系数菱形多项式族

$$P=\left\{p(s)=\sum_{i=0}^{n}a_is^i,\left|\sum_{i=0}^{n}\left|a_i-a_i^0\right|\leqslant r\right.\right\},$$

其中

$$p_0(s)\triangleq\sum_{i=0}^{n}a_i^0s^i$$

为标称多项式,全族的稳定性(连续时间意义下)可由下列八个顶点多项式的稳定性所保证:

$$p_0(s)\pm r,\qquad p_0(s)\pm rs,$$
$$p_0(s)\pm rs^{n-1},\qquad p_0(s)\pm rs^n.$$

这一定理有两个重要特点:

1. 所需检验的顶点多项式数目与多项式族的阶次无关.

2. 只适用于连续时间意义下系统的稳定性判定.

哈里托诺夫区域(Kharitonov region) 对于区间多项式族,哈里托诺夫型结论成立的稳定区域. 设 $D\subset\mathrm{C}$ 为复平面上的一个开区域,如果多项式族

$$P=\{p(s,q),q\in Q\}$$

中的任一多项式的根都在 D 内,则称 P 是鲁棒 D 稳定的. 如果对于任一不降次的区间多项式族

$$P=\left\{p(s)=\sum_{i=0}^{n}a_is^i,a_i\in[a_i^-,a_i^+]\right\},$$

其 D 稳定性均能由其顶点多项式族

$$V=\left\{p(s)=\sum_{i=0}^{n}a_is^i,a_i\in[a_i^-,a_i^+]\right\}$$

的 D 稳定性所保证, 则称 D 为一弱哈里托诺夫区域. 设 $D\subseteq\mathrm{C}$,则 D 是一个弱哈里托诺夫区域,当 D 和 $1/D=\{z:zd=1,\exists\ d\in D\}$ 都是凸的.

这一结论表明:当 D 与 $1/D$ 均为凸集时,区间多项式族的 D 稳定性可由其顶点多项式族的 D 稳定性所保证.

凸方向(convex direction) 稳定多项式空间的一个特定方向. 设 $g(s)$ 为一首项系数为 1 的多项式,如果对于任意使得 $f(s)+g(s)$ 稳定且 $\deg(f(s)+\lambda g(s))=n(\forall\lambda\in[0,1])$ 的 n 次稳定多项式 $f(s)$,都有 $f(s)+\lambda g(s)$ 稳定,$\forall\lambda\in[0,1]$,则称 $g(s)$ 为 n 次稳定多项式空间的一个凸方向.

定理:多项式 $g(s)$ 是 n 次实系数赫尔维茨稳定多项式空间的一个凸方向的充分必要条件是

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\arg(g(\mathrm{j}\omega))\leqslant\left|\frac{\sin 2\arg(g(\mathrm{j}\omega))}{2\omega}\right|$$
$$(\forall\omega>0,g(\mathrm{j}\omega)\neq0),$$

这里 arg(·)表示相角.

上述结果将以往的关于多项式族鲁棒稳定性的有限检验结果统一在一个框架下,对于离散时间意义下的舒尔稳定性可通过双线性变换得到类似的结论.

棱边定理(edge theorem) 凸多面体多项式族(即在多项式系数空间中形成凸多面体的一族多项式)稳定性的一维突出棱边检验定理. 对于 n 次实系数凸多面体多项式族

$$P=\left\{p(s)=\sum_{i=0}^{n}\lambda_ip_i(s)\,|\,\lambda_i\geqslant0,\sum_{i=0}^{n}\lambda_i=1\right\},$$

其中 $p_i(s)(i=1,2,\cdots,m)$ 为 m 个顶点多项式. 全族的稳定性(包括连续、离散时间意义下的稳定性,稳定区域可为复平面上相当一般的区域)可由形如

$$\lambda p_i(s)+(1-\lambda)p_j(s),\ \lambda\in[0,1]$$

的所有一维突出棱边多项式的稳定性所保证. 这一定理有三个重要特点:

1. 已有反例说明两个顶点多项式的稳定性一般不能保证其凸组合(棱边)多项式的稳定性,因此棱边定理是所能期望的最好结果.

2. 稳定区域可以是复平面上相当一般的区域,因此棱边定理为性能鲁棒性的研究提供了有力工具.

3. 一维棱边的稳定性检验已有有效的算法,因此,棱边检验的条件是容易检验的.

边界检验(boundary verification) 多项式参数空间或系数空间中几何体稳定性的检验方法. 其稳定性可由其相对边界的稳定性所保证. 边界检验一般要求参数到多项式系数的映射为仿射映射. 关于边界检验的主要结果有边界定理和原象定理.

边界定理是关于多项式参数空间中几何体的稳定性的相对边界检验定理. 对 n 次实系数多项式参数空间中的任一几何体,当参数空间与系数空间以仿射映射相联系时,该几何体所对应的多项式族的稳定性可由其相对边界(相对于仿射包的边界)所对应的多项式族的稳定性所保证.

原象定理是关于多项式系数空间中几何体的稳定性的另一个判别定理. 对于多项式族 P 和稳定区域 $D,\forall z\in\partial D$ 在值映射 V 下,P 在复平面上的象(即值集)$V(P)$的边界为$\partial V(P)$,若$\forall c\in\partial V(P)$,在 c 的原象 $V^{-1}(c)$中均存在一个稳定的多项式,则多项式族 P 是稳定的. 这两个定理都可以是针对一般意义下 D 稳定性的,并且由此出发可以导出多项式族稳定性的一系列重要结果.

值映射(value mapping) 研究多项式族稳定性用到的数学概念. 对于多项式 $p(s)$和复数 z,值映射 V_z 定义为

$$V_zp\triangleq p(z)\in\mathrm{C}.$$

多项式系数空间中的一个几何体在值映射下为复平面上的一个几何图形. 对于多项式族 P 和复数 z,$P(z)$称为值集. 值映射与值集是多项式族鲁棒稳定性分析的有力工具,它与剔零原理一起构成多项式

族鲁棒稳定性研究的基本方法.多项式族在值映射下象(即值集)的几何性质在鲁棒稳定性检验中有重要作用.

值集(value set)　见“值映射”.

剔零原理(zero exclusion principle)　现代鲁棒控制研究中的重要原理.又称排零原理、除零原理.对于 n 次多项式族 $P(s)$ 和稳定区域 D,$P(s)$ 稳定,当且仅当:

1. 存在 $p_0(s)\in P(s)$ 是 D 稳定的.
2. $0\notin P(z)\triangleq\{p(z)\mid p(s)\in P(s)\},\forall z\in\partial D$,其中 ∂D 表示 D 的边界.

上述条件 2 表明:在 $z\in\partial D$ 时多项式族的值集将坐标原点(即零点)排除在外,因而称排零原理.

剔零原理可以追溯到 20 世纪 20 年代,是现代鲁棒控制研究中关于多项式族稳定性的理论基础,是多项式根与系数连续依赖性的具体表现.

映射定理(mapping theorem)　多仿射映射下多项式族的值集性质的重要定理.设

$$f(s,\gamma)=s^n+\sum_{i=1}^{n}a_i(\gamma)s^{n-i},$$
$$\gamma=(\gamma_1,\gamma_2,\cdots,\gamma_m)^{\mathrm{T}}\in\Gamma$$
$$\triangleq\{\gamma\mid 0\leqslant\gamma_i\leqslant 1,i=1,2,\cdots,m\},$$

其中 $a_i(\gamma)(i=1,2,\cdots,m)$ 为 $\gamma_1,\gamma_2,\cdots,\gamma_m$ 的多仿射函数.记

$$\Gamma_0\triangleq\{\gamma\mid\gamma_i=0,1;i=1,2,\cdots,m\},$$

则有

$$\mathrm{conv}f(\mathrm{j}\omega,\Gamma)=\mathrm{conv}f(\mathrm{j}\omega,\Gamma_0),$$

这里 $\omega\in\mathrm{R}$,j 表示虚数单位,conv(·)表示 R^2 上集合的凸包.该定理是研究多仿射映射下多项式族的稳定性的重要工具之一.

主点(principal point)　多仿射映射下多项式族稳定性检验的一个重要概念.设 z_1,z_2 为两个非零复数,如果 $Im(z_2/z_1)>0$,则记 $z_1\prec z_2$;如果 $Im(z_2/z_1)\geqslant 0$,则记 $z_1\preceq z_2$;如果 $Im(z_2/z_1)=0$,则记 $z_1\asymp z_2$.

考虑多仿射映射 $f:B^m\to\mathrm{C}$,其中 C 为复平面,$B^m=\{x\mid x\in\mathrm{R}^m,a_k\leqslant x_k\leqslant b_k,k=1,2,\cdots,m\}$,并且 $f(x_1,x_2,\cdots,x_m)$ 为 $x_1,x_2,\cdots,x_m$ 的多仿射函数,即

$$f(x_1,x_2,\cdots,x_k+\gamma,\cdots,x_m)$$
$$-f(x_1,x_2,\cdots,x_k,\cdots,x_m)$$
$$=\gamma\frac{\partial f}{\partial x_k}(x)\triangleq\gamma G_k(x)$$
$$(k=1,2,\cdots,m).$$

对于 $x\in B^m$,定义下面三个指标集:

1. $I_a=\{i\mid a_i=x_i\}$.
2. $I_0=\{i\mid a_i<x_i<b_i\}$.
3. $I_b=\{i\mid x_i=b_i\}$.

对 $x\in B^m$,如果存在非零复数 G,使得

$$G_i(x)\asymp G,\ \forall i\in I_0;\ G_i(x)\preceq G\preceq G_k(x),$$
$$\forall G_i\neq 0,\ i\in I_a;\ G_k\neq 0,\ k\in I_b,$$

则称 x 为一个主点.设 $z\in\partial f(B^m)$,则一定存在一个主点 x,使得 $z=f(x)$.对于一个多项式族

$$P(s)=\left\{p(s,x)\mid p(s,x)=\sum_{i=0}^{n}a_i(x)s^i,x\in B^m\right\},$$

其中 $a_i(x)(i=0,1,\cdots,n)$ 是 R^m 到 R 的多仿射函数,其稳定性可通过寻找其值映射下的主点集来判定.如果主点集都在其突出棱边上,则棱边检验结果成立.一般情况下棱边检验不能保证全族的稳定性.对于一些特别的系统,可找到一些主边和主节,这些主边和主节的稳定性可保证全族的稳定性.

鲁棒严格正实性(robust strictly positive realness)　具有参数摄动的控制对象族的严格正实性.设 $P(s)=\{p(s,q)=N(s,q)/D(s,q),q\in Q\}$ 是一个不确定对象族,如果对于任意的 $q\in Q$,$p(s,q)$ 都是严格正实的,则称 $P(s)$ 是鲁棒严格正实的.

当对象族的分子 $N(s,q)$ 与分母 $D(s,q)$ 相互独立,且参数到系数间为仿射映射时,全族的鲁棒严格正实性具有凸性;如果分子、分母分别在两个相互独立的多面体中变化,则全族的严格正实性可由其顶点对象的严格正实性所保证.设

$$P(s)=\left\{p(s,q,t)=\frac{N(s,q)}{D(s,t)}\,\middle|\,q\in Q,t\in T\right\},$$
$$N(s,q)=\sum_{i=0}^{m}a_i(q)s^i,$$
$$D(s,t)=\sum_{i=0}^{n}b_i(t)s^i,$$

其中 $m\leqslant n,q=(q_1,q_2,\cdots,q_{l_1})^{\mathrm{T}},t=(t_1,t_2,\cdots,t_{l_2})^{\mathrm{T}}$ 分别为 l_1 维、l_2 维实向量,$Q\subset\mathrm{R}^{l_1},T\subset\mathrm{R}^{l_2}$ 均为凸多面体,记 $V(Q),V(T)$ 为对应的顶点集,$b_n(t)\neq 0,\forall t\in T$.当 $a_i(q),b_j(t)(i=0,1,\cdots,m;j=0,1,\cdots,n)$ 分别为 $q_1,q_2,\cdots,q_{l_1}$ 和 $t_1,t_2,\cdots,t_{l_2}$ 的仿射函数时,$P(s)$ 的严格正实性可由下面的顶点对象族的严格正实性所保证:$V(s)=\{p(s,q,t)\mid t\in V(T),q\in V(Q)\}$.

当 Q,T 为超矩形,$a_i(q),b_j(t)$ 为多仿射函数时,$P(s)$ 的严格正实性仍能由 $V(s)$ 的严格正实性所保证.特别地,对于区间对象族,即

$$P(s)=\left\{p(s,q,t)=\frac{N(s,q)}{D(s,t)}\,\middle|\,q\in Q,t\in T\right\},$$
$$N(s,q)=\sum_{i=0}^{m}q_is^i,\ D(s,t)=\sum_{i=0}^{n}t_is^i,$$
$$Q=\{q=(q_0,q_1,\cdots,q_m)^{\mathrm{T}}\mid q_i^-\leqslant q_i\leqslant q_i^+,$$
$$i=0,1,\cdots,m\},$$
$$T=\{t=(t_0,t_1,\cdots,t_n)^{\mathrm{T}}\mid t_i^-\leqslant t_i\leqslant t_i^+,$$
$$i=0,1,\cdots,n\},$$

其中 $t_n \neq 0, \forall t \in T$. 则 $P(s)$的严格正实性(SPR)或KSPR(即 $K+P(s)$是鲁棒 SPR 的,$K \geqslant 0$)性质可由下面的 16 个顶点对象族的 SPR 或 KSPR 性质所保证：

$$p_{i,k}(s)=\frac{N_i(s)}{D_k(s)} \quad (i,k=1,2,3,4),$$

其中 $N_i(s), K_k(s)(i,k=1,2,3,4)$分别为对应区间多项式 $N(s,q)$和 $D(s,t)$的四个哈里托诺夫多项式(参见"哈里托诺夫定理"). 上述顶点对象通常称为哈里托诺夫对象. 鲁棒严格正实性的研究背景是自适应系统的算法收敛性分析和非线性系统的鲁棒绝对稳定性.

非线性参数化(nonlinear parameterization) 多项式族系数非线性地依赖于不确定参数的简称. 对这类多项式族的研究是参数不确定系统鲁棒稳定性研究的方向之一. 该问题的提法如下：设系统不确定参数的变化范围是参数空间 R^m 中的有界闭集 $Q \subset \mathrm{R}^m$. 系统多项式

$$P(s) = \sum_{i=0}^{n} a_i s^i$$

的系数 $a_i(i=0,1,\cdots,n)$是不确定参数 $q \in Q$ 的非线性函数,即 $a_i = a_i(q)(i=0,1,\cdots,n)$是非线性的. 当不确定参数 q 在 Q 中变化时,得到的多项式族称为非线性参数化多项式族,记为

$$P(s,Q) = \left\{P(s,q) = \sum_{i=0}^{n} a_i(q) s^i, q \in Q\right\}.$$

若多项式族 $P(s,Q)$中的每一多项式的根都位于复平面 C 中给定的稳定域 $D \subset \mathrm{C}$ 中时,非线性参数化多项式族 $P(s,Q)$称为鲁棒 D 稳定的. 该方向研究的主要问题是如何给出一种可操作的检验方法来确定多项式族 $P(s,Q)$的鲁棒 D 稳定性. 目前,这方面有价值的研究结果还不多见.

鲁棒镇定(robust stabilization) 一种对系统稳定性的控制. 指用一个确定的控制器来控制不确定对象,使所构成的闭环不确定系统具有要求的稳定性. 参数不确定系统的鲁棒镇定问题的提法如下：给定对象族

$$G(s;Q,R) = N(s,Q)/D(s,R),$$

其中 $N(s,Q), D(s,Q)$均为多项式族,Q,R 分别为分子、分母多项式族的不确定参数集. 若存在控制器 $C(s)=n(s)/d(s)$(其中 $n(s), d(s)$分别为控制器的分子、分母多项式),使得闭环系统的特征多项式族

$$P(s;Q,R) = N(s,Q)n(s) + D(s,R)d(s)$$

是鲁棒 D 稳定的,则称控制器 $C(s)$鲁棒镇定对象族 $G(s;Q,R)$. 该方向研究的问题是：对给定的对象族 $G(s;Q,R)$,是否存在一个确定的控制器鲁棒镇定对象族 $G(s;Q,R)$,以及怎样求取这样的控制器. 目前,这方面的研究是人们所关心的主要问题之一.

设不确定对象族的分子、分母多项式族为区间多项式族：

$$g(s;q,r) = \frac{\sum_{i=0}^{m}[q_i^-, q_i^+]s^i}{\sum_{i=0}^{n}[r_i^-, r_i^+]s^i},$$

对于分子、分母多项式族分别有四个哈里托诺夫顶点多项式(参见"哈里托诺夫定理"),以及由相同偶部或奇部的哈里托诺夫顶点多项式的凸组合所构成的哈里托诺夫边. 固定分子(或分母)多项式为某一哈里托诺夫顶点多项式；分母(或分子)多项式取哈里托诺夫边,这些边所构成的对象族称为 CB 棱边,这样的棱边对象族共有 32 个.

盒子定理(box theorem) 检验控制器使区间对象族赫尔维茨稳定的一个充分必要条件. 控制器鲁棒赫尔维茨镇定区间对象族的充分必要条件是它同时镇定 32 个事先给定的棱边对象族. 盒子定理是由美国的查普利特(Chapellat, H)和巴特切亚(Bhattacharyya, S. P.)于 1989 年给出的.

顶点镇定定理(vertex stabilization theorem) 检验一阶控制器使区间对象族赫尔维茨稳定的一个充分必要条件. 一阶控制器能鲁棒赫尔维茨镇定区间对象族的充分必要条件是它同时镇定 16 个顶点对象,这 16 个顶点对象是由分子、分母多项式分别取相应的哈里托诺夫顶点多项式而构成的.

顶点镇定定理是由美国的巴米希(Barmish, B. R.)等人于 1992 年给出的. 参数不确定系统鲁棒镇定的研究目前还只处于分析阶段,这主要是由于多对象(两个以上)同时镇定问题尚未解决.

摄动界(perturbation bound (or margin)) 表征系统稳定裕度大小的一个指标. 表示稳定的标称系统,当其参数在给定的摄动模式下摄动时,系统稳定裕度大小的一个指标. 设稳定的标称系统的特征多项式为

$$P(s,q^0) = \sum_{i=0}^{n} a_i(q^0) s^i,$$

其中多项式系数是系统参数 $q \in \mathrm{R}^m$ 的函数. 摄动模式可表示为定义在参数空间 R^m 上的拟凸函数 $f:\mathrm{R}^m \to \mathrm{R}$,并且满足

$$\min_{q \in \mathrm{R}^m} f(q) = f(q^0) = 0.$$

带有摄动半径 $\varepsilon > 0$ 的摄动域定义为

$$Q(q^0,\varepsilon) = \{q \in \mathrm{R}^m : f(q) \leqslant \varepsilon\}.$$

从而可定义参数化多项式族：

$$P(s;Q(q^0,\varepsilon)) = \{P(s,q) = \sum_{i=0}^{n} a_i(q) s^i, \ q \in Q(q^0,\varepsilon)\}.$$

在该摄动模式下的 D 稳定摄动界为

$$r = \min_{\varepsilon > 0}\{\varepsilon: P(s, Q(q^0, \varepsilon)) \text{ 不是 } D \text{ 稳定的}\}.$$

摄动界的研究，根据稳定域 D、系统参数化的形式 $a_i(q)(i=0,1,\cdots,n)$ 以及摄动模式 $f(q)$ 的不同，又可进行更细的分类. 目前主要的研究结果限于赫尔维茨稳定及舒尔稳定（稳定域 D 为单位圆内部）系统在系数空间以及仿射参数化情况下参数空间中的范数摄动模式，即

$$f(q) = \left\{ \sum_{i=0}^{m-1} \left(\frac{q_i - q_i^0}{\alpha_i} \right)^p \right\}^{1/p},$$

其中 $\alpha_i(i=0,1,\cdots,m-1)$ 是正常数.

当 $p=1$ 时，称为菱形摄动模式；$p=2$ 时，称为球形摄动模式；$p=\infty$ 时，称为区间摄动模式. 其相应的摄动界估计式均已得到.

区间矩阵（interval matrix） 矩阵的元素在给定区间上摄动时所对应的一类矩阵族. 给定如下的实区间矩阵族：

$$A' = \{A = [a_{ij}]_{n\times n}: a_{ij}^- \leqslant a_{ij} \leqslant a_{ij}^+; i,j = 1,2,\cdots,n\}.$$

若对所有的矩阵 $A \in A'$，其特征值都位于复平面中给定的稳定域 D 中，则称区间矩阵族 A' 是 D 稳定的. 目前的研究主要限于赫尔维茨稳定及舒尔稳定. 区间矩阵的元取为相应区间的顶点所对应的矩阵称为顶点矩阵. 主要结果有：当 A' 为 2×2 时，顶点矩阵的赫尔维茨稳定性等价于 A' 的赫尔维茨稳定性；当 A' 为 3×3，且对角线元素固定不变时，即

$$a_{ii}^- = a_{ii}^+ \quad (i = 1,2,3),$$

顶点矩阵的赫尔维茨稳定性等价于 A' 的赫尔维茨稳定性；对于 3×3 以上的区间矩阵族 A'，当族中每一矩阵为对称矩阵时，顶点矩阵的赫尔维茨稳定性等价于 A' 的赫尔维茨稳定性；当族中每一矩阵为正规矩阵时，顶点矩阵的赫尔维茨稳定性等价于 A' 的赫尔维茨稳定性. 对于 3×3 以上的区间矩阵族的稳定性的有限检验，有价值的结果还不多见. 现已证明：对于一般的矩阵凸多面体：

$$A' = \left\{A = \sum_{i=1}^{l} \lambda_i A_i : \lambda_i \geqslant 0, \sum_{i=1}^{l} \lambda_i = 1, A_i \text{ 固定}\right\},$$

当 A' 为 3×3 以上时，必须检验所有的 $(2n-4)$ 维边界的赫尔维茨稳定性，当所有的 $(2n-5)$ 维边界都稳定时，A' 中仍可能存在不稳定矩阵. 目前该方向的研究是在某些合理的假设下，寻找区间矩阵族稳定性的某些充分条件，或将问题转化为数学规划来求解.

H_∞控制理论（H_∞ control theory） 一种在保证系统稳定前提下的最优控制理论. 它源于 20 世纪 80 年代初，是由加拿大学者詹姆斯（Zames, G.）首先提出，后经多伊尔（Doyle, J. C.）、弗朗希斯（Francis, B. A.）、格朗沃尔（Glover, K.）等人进一步完善和发展，至今已成为一门较成熟的学科分支. 特别是在线性定常系统的范围内，H_∞ 控制已形成了一套完整的理论体系，并开发了用于系统设计的软件工具箱. 在时变系统、非线性系统以及分布参数系统领域，H_∞ 控制理论也得到了不同程度的发展.

就定常线性系统而言，H_∞ 控制理论主要是讨论控制器的综合问题，要求在保证闭环系统稳定的前提下，使得系统的外部干扰对调节输出造成的影响达到最小. 而对应系统的传递函数矩阵的 H_∞ 范数恰恰是对由外部干扰对调节输出所造成影响大小的一种度量，这就是人们为什么致力于 H_∞ 控制理论研究的基本出发点. 此外，H_∞ 控制理论也是研究鲁棒控制的一种重要手段. H_∞ 控制理论有两种研究途径：频域法和状态空间方法. 从频域法出发，H_∞ 控制问题最终将转化成一个模型匹配问题，然后采取不同的处理手段解这个模型匹配问题来达到控制器的综合目标. 从状态空间出发，H_∞ 控制问题将导致解两个代数黎卡提方程.

H_2 空间（H_2-space） 亦称哈代 2 空间. 是一种函数空间. 它是由在右半开复平面内解析且范数有界的复变量向量值函数的全体组成的函数空间. 一般定义为：$H_2 \triangleq \{f: f(s)$ 在 $\mathrm{Re}\, s > 0$ 解析，且

$$\|f\|_2 = \left[\sup_{\alpha>0} \frac{1}{2\pi}\int_{-\infty}^{\infty} f^*(\alpha + \mathrm{j}\omega) f(\alpha + \mathrm{j}\omega)\mathrm{d}\omega\right]^{1/2} < \infty\},$$

其中"$*$"表示向量或矩阵的共轭转置，$\mathrm{Re}\{\cdot\}$ 表示复数的实部，"j"表示虚数单位.

对于给定的复变量向量值函数 $f(\mathrm{j}\omega)$，定义

$$\|f\|_2 = \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} f^*(\mathrm{j}\omega) f(\mathrm{j}\omega)\mathrm{d}\omega\right]^{1/2}.$$

称 $\|f\|_2$ 为 $f(s)$ 的 L_2 范数. 称集合

$$L_2 \triangleq \{f: \|f\|_2 < \infty\}$$

为 L_2 空间. L_2 空间是一个希尔伯特空间，它与时域 $L_2(-\infty, \infty)$ 空间同构. 对于每个

$$f \in H_2, f_b(\mathrm{j}\omega) = \lim_{\alpha\to 0} f(\alpha + \mathrm{j}\omega),$$

对几乎所有的 ω 存在，通常称 $f_b(\mathrm{j}\omega)$ 为 $f(\alpha+\mathrm{j}\omega)$ 的边界函数，并且 $\|f_b\|_2 = \|f\|_2$. 因此由 $f \in H_2$ 到 $f_b \in L_2$ 的映像是线性、内射和保范的. 人们可把 H_2 空间看成 L_2 空间的一个子空间，它与时域 $L_2[0, \infty)$ 空间同构. 若令

$$H_2^\perp \triangleq \{f: f(-s) \in H_2\},$$

那么 $H_2^\perp$ 与时域 $L_2[0, \infty)$ 空间同构，且 $L_2 = H_2 \oplus H_2^\perp$，$H_2^\perp$ 为 H_2 的正交补空间.

哈代2 空间（Hardy 2-space） 即"H_2 空间".

H_∞ 空间（H_∞-space） 亦称哈代空间. 是一种函数空间. 它是由在右半开复平面内解析的矩阵值函数的全体组成的函数空间，其定义为

$$H_\infty \triangleq \{G(s): G(s)\text{在 } \mathrm{Re}(s)>0 \text{ 解析,且} \sup_{\mathrm{Re}(s)>0} \bar{\sigma}(G(s))<\infty\},$$

其中

$$\|G\|_\infty = \sup_{\mathrm{Re}(s)>0} \bar{\sigma}(G(s))$$

为 $G(s)$的 H_∞范数,$\bar{\sigma}(\cdot)$表示矩阵的最大奇异值.记 $RH_\infty \subset H_\infty$为在右半闭复平面内解析的真实有理函数矩阵的全体.对于给定的 $m\times n$ 阶复的矩阵值函数 $G(\mathrm{j}\omega)$,定义

$$\|G\|_\infty = \mathrm{ess}\sup_{\omega} \bar{\sigma}(G(\mathrm{j}\omega))$$

为 $G(\mathrm{j}\omega)$的 L_∞ 范数,其中"ess sup"表示本质上确界(即除去 ω 的一个零测子集之后的上确界).称集合

$$L_\infty \triangleq \{G(s) \mid \|G(s)\|_\infty<\infty\}$$

为 L_∞空间.记 $RL_\infty \subset L_\infty$为真实有理且在虚轴上无极点的矩阵的全体.若 $G(s)\in RH_\infty$,那么可以把 $G(s)$看做是某个定常线性系统的传递函数矩阵,则对所有的 $u(s)\in L_2$ 或 H_2,都有 $G(s)u(s)\in L_2$ 或 H_2,并且

$$\|G\|_\infty = \sup_{\|u\|_2=1} \|G(s)u(s)\|_2,$$

即 $G(s)$的 H_∞范数和它的诱导算子范数相等.上式右端 $\|\cdot\|_2$ 是对应空间 L_2 或 H_2 范数.

哈代空间(Hardy space) 即"H_∞空间".

H_∞控制的标准问题(standard problem for H_∞ control) H_∞控制理论所研究的基本问题.考虑如图所示的 H_∞控制问题的标准结构,其广义受控对象的方程为

$$z(s)=P_{11}(s)w(s)+P_{12}(s)u(s),$$
$$y(s)=P_{21}(s)w(s)+P_{22}(s)u(s),$$

其中 $z(s)$,$y(s)$,$u(s)$,$w(s)$分别为广义受控对象的 p 维调节输出、m 维量测输出、r 维控制输入和 q 维干扰输入信号.$P_{ij}(s)$ $(i,j=1,2)$分别为具有相应维数的传递函数矩阵,它们都是真有理分式矩阵.通常令

广义受控对象H_∞控制问题的标准结构

$$P(s)=\begin{pmatrix} P_{11}(s) & P_{12}(s) \\ P_{21}(s) & P_{22}(s) \end{pmatrix}$$

表示广义受控对象的传递函数矩阵.这个广义受控对象的状态空间实现为

$$\dot{x}(t)=Ax(t)+B_1w(t)+B_2u(t),$$
$$z(t)=C_1x(t)+D_{11}w(t)+D_{12}u(t),$$
$$y(t)=C_2x(t)+D_{21}w(t)+D_{22}u(t),$$

其中 $x(\cdot)$为 n 维状态向量,A,B_1,B_2,C_1,C_2,D_{ij} $(i,j=1,2)$分别为具有相应维数的常值矩阵.采用线性系统理论的表示方法可记为

$$P(s)=\left[\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array}\right].$$

这样 $P_{ij}(s)=C_i(sI-A)^{-1}B_j+D_{ij}$ $(i,j=1,2)$,其中 I 表示单位矩阵,其维数视上下文而定.

所谓 H_∞控制的标准问题是说,对于给定的广义受控对象 $P(s)$,求一个反馈控制器 $K(s)$,使得当取 $u(s)=K(s)y(s)$时闭环系统是内部稳定的,同时由外部干扰输入 $w(s)$到系统的调节输出$z(s)$之间的传递函数矩阵的 H_∞范数达到极小.如果这样的反馈控制器$K(s)$存在,那么就称这个 $K(s)$为 H_∞控制问题的最优解.如果存在一个反馈控制器 $K(s)$,使得当取$u(s)=K(s)y(s)$时闭环系统是内部稳定的,并且由外部干扰输入 $w(s)$到系统的调节输出 $z(s)$之间的传递函数矩阵的 H_∞范数小于(或小于等于)某个事先给定的正数 γ,那么就称这个 $K(s)$为 H_∞控制问题的次优解.

线性分式变换(linear fractional transformation) 复变量(矩阵变量)之间的一种变换.在控制理论中可用于描述动态系统,是一种描述系统的基本构架.它既可描述系统的传递函数、系统的不确定性等系统的基本特征,又可描述具有特定性质的镇定控制器的一般结构.因此它是现代控制理论中的一种重要的研究工具.假设 M 是一个$(p_1+p_2)\times(q_1+q_2)$阶的复矩阵,并按分块形式表示成

$$M=\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

其中 $M_{11},M_{12},M_{21},M_{22}$分别为 $p_1\times q_1,p_1\times q_2,p_2\times q_1,p_2\times q_2$ 阶的复矩阵.再令 $\mathscr{D}_1\subset C^{q_1\times p_1}$,$\mathscr{D}_2\subset C^{q_2\times p_2}$都是开集,这里 $C^{q_i\times p_i}$表示 $q_i\times p_i$ 阶复矩阵的全体组成的集合,$i=1,2$.定义映像

$$\mathscr{F}_l(M,\cdot): \mathscr{D}_2 \mapsto C^{p_1\times q_1},$$
$$\mathscr{F}_u(M,\cdot): \mathscr{D}_1 \mapsto C^{p_2\times q_2},$$

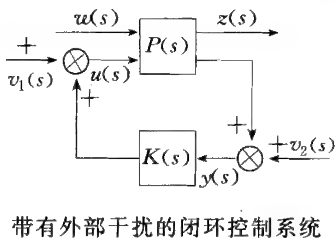
其中

$$\mathscr{F}_l(M,\Delta_l)\triangleq M_{11}+M_{12}\Delta_l(I-M_{22}\Delta_l)^{-1}M_{21},$$
$$\mathscr{F}_u(M,\Delta_u)\triangleq M_{22}+M_{21}\Delta_u(I-M_{11}\Delta_u)^{-1}M_{12}$$
$$(\Delta_l\in\mathscr{D}_2,\ \Delta_u\in\mathscr{D}_1).$$

如果在这些映像中的逆矩阵存在,那么分别把 $\mathscr{F}_l(M,\Delta_l)$和 $\mathscr{F}_u(M,\Delta_u)$称为相对参数矩阵 Δ_l 和 Δ_u 的下线性分式变换和上线性分式变换,矩阵 M 称为该变换的系数矩阵.在给定广义受控对象 $P(s)$后,可以把由控制器 $K(s)$形成的闭环系统由 $w(s)$到 $z(s)$的传递函数矩阵写成以 $P(s)$为系统矩阵、$K(s)$为参数矩阵的下线性分式变换形式 $\mathscr{F}_l(P(s),K(s))$.在近代有关 H_∞控制的文献中,常用它表示闭环系统的传递函数矩阵.

适定性(well-posedness) 保证闭环系统能正

常工作的基本特性.前者是指控制系统的物理可实现性,后者是指反馈控制系统中由任何输入到任何输出之间的传递函数矩阵都是稳定的.已知广义受控对象和控制器分别为



带有外部干扰的闭环控制系统

$$z(s)=P_{11}(s)w(s)+P_{12}(s)u(s),$$
$$y(s)=P_{21}(s)w(s)+P_{22}u(s)+v_2(s),$$
$$u(s)=K(s)y(s)+v_1(s),$$

其中 $v_1(s),v_2(s)$ 分别为具有适当维数的外部干扰信号,其余符号同前.如果由 $w(s),v_1(s),v_2(s)$ 到 $z(s),u(s),y(s)$ 之间的九个传递函数矩阵都是真有理分式矩阵,那么就称闭环系统是适定的.如果这九个传递函数矩阵都是稳定的,那么就称闭环系统是内部稳定的,同时称控制器 $K(s)$ 镇定广义受控对象 $P(s)$,即 $K(s)$ 为 $P(s)$ 的镇定控制器.

内部稳定性(internal stability) 见“适定性”.

在 RH_∞ 上的互质分解(coprime factorization over RH_∞) 正则实有理函数矩阵的一种分解.RH_∞ 中的两个矩阵 $F(s)$ 和 $G(s)$ 称为是右互质的,如果它们有相同的列数,并且存在矩阵 $X(s),Y(s)\in RH_\infty$,使得

$$X(s)F(s)+Y(s)G(s)=I.$$

这等价于矩阵

$$\begin{pmatrix}F(s)\\G(s)\end{pmatrix}$$

在 RH_∞ 中是左可逆的.同样,称 $F(s)$ 和 $G(s)$ 是左互质的,如果它们有相同的行数,并且存在矩阵 $X(s)$,$Y(s)\in RH_\infty$,使得

$$F(s)X(s)+G(s)Y(s)=I.$$

这等价于矩阵 $[F(s),G(s)]$ 在 RH_∞ 中是右可逆的.现在令 $G(s)$ 是一个真实有理矩阵.如果存在 RH_∞ 中的右互质矩阵 $N(s)$ 和 $M(s)$,使得

$$G(s)=N(s)M^{-1}(s),$$

那么就称 $N(s)M^{-1}(s)$ 为 $G(s)$ 的右互质分解.同样,如果存在 RH_∞ 中的左互质矩阵 $\tilde{N}(s)$ 和 $\tilde{M}(s)$,使得

$$G(s)=\tilde{M}^{-1}(s)\tilde{N}(s),$$

那么就称 $\tilde{M}^{-1}(s)\tilde{N}(s)$ 为 $G(s)$ 的左互质分解.对于每个真实有理矩阵 $G(s)$,都存在 RH_∞ 中的八个矩阵 $M(s),N(s),\tilde{M}(s),\tilde{N}(s),X(s),Y(s),\tilde{X}(s),\tilde{Y}(s)$,使得

$$G(s)=N(s)M^{-1}(s)=\tilde{M}^{-1}(s)\tilde{N}(s),$$

$$\begin{pmatrix}\tilde{X}(s) & -\tilde{Y}(s)\\-\tilde{N}(s) & \tilde{M}(s)\end{pmatrix}\begin{pmatrix}M(s) & Y(s)\\N(s) & X(s)\end{pmatrix}=\begin{pmatrix}I & 0\\0 & I\end{pmatrix}.$$

如果 $G(s)$ 的状态空间实现为

$$G(s)=\left(\begin{array}{c|c}A & B\\\hline C & D\end{array}\right),$$

并且 (A,B) 能稳,(C,A) 能检测,则有常数矩阵 F,H 存在,使得

$$M(s)\triangleq\left(\begin{array}{c|c}A_F & B\\\hline F & I\end{array}\right),\qquad N(s)\triangleq\left(\begin{array}{c|c}A_F & B\\\hline C_F & D\end{array}\right),$$

$$\tilde{M}(s)\triangleq\left(\begin{array}{c|c}A_H & H\\\hline C & I\end{array}\right),\qquad \tilde{N}(s)\triangleq\left(\begin{array}{c|c}A_H & B_H\\\hline C & D\end{array}\right),$$

$$X(s)\triangleq\left(\begin{array}{c|c}A_F & -H\\\hline C_F & I\end{array}\right),\qquad Y(s)\triangleq\left(\begin{array}{c|c}A_F & -H\\\hline F & 0\end{array}\right),$$

$$\tilde{X}(s)\triangleq\left(\begin{array}{c|c}A_H & -B_H\\\hline F & I\end{array}\right),\qquad \tilde{Y}(s)\triangleq\left(\begin{array}{c|c}A_H & -H\\\hline F & 0\end{array}\right),$$

其中 $A_F\triangleq A+BF$ 和 $A_H\triangleq A+HC$ 都是稳定矩阵,而 $C_F\triangleq C+DF$,$B_H\triangleq B+HD$.

控制器的参数化(parameterization of controller) 带有自由参数、具有某种性质的控制器的一般表示.已知广义受控对象 $P(s)$、控制器 $K(s)$ 镇定 $P(s)$ 的充分必要条件是 $K(s)$ 镇定 $P_{22}(s)$.从控制器的设计角度,只需研究 $P_{22}(s)$ 的镇定即可,为书写方便,略去下标.令 $P(s)$ 的左、右互质分解分别记为 $(\tilde{M}(s),\tilde{N}(s))$ 和 $(M(s),N(s))$,则镇定 $P(s)$ 的所有控制器 $K(s)$ 的集合可表示成如下的参数化公式:

$$K(s)=(Y(s)-M(s)Q(s))(X(s)-N(s)Q(s))^{-1}$$
$$=(\tilde{X}(s)-Q(s)\tilde{N}(s))^{-1}(\tilde{Y}(s)-Q(s)\tilde{M}(s)),$$
$$\det(X(s)-N(s)Q(s))\neq 0,$$
$$\det(\tilde{X}(s)-Q(s)\tilde{N}(s))\neq 0,$$

其中 $Q(s)\in RH_\infty$ 是任意的,$\det(\cdot)$ 表示矩阵的行列式.由此可见,给定广义受控对象 $P(s)$ 以后,只要对 $P_{22}(s)$ 进行左(或右)互质分解,进而求得 $\tilde{X}(s)$ 和 $\tilde{Y}(s)$(或 $X(s)$ 和 $Y(s)$),就可得到镇定控制器的一般表达式,然后选择适当的自由参数 $Q(s)\in RH_\infty$,使闭环系统满足设计要求.这就是控制设计的一般步骤.关于 $\tilde{X}(s)$ 和 $\tilde{Y}(s)$(或 $X(s)$ 和 $Y(s)$)与 $\tilde{M}(s)$ 和 $\tilde{N}(s)$(或 $M(s$ 和 $N(s)$)之间的关系参见“在 RH_∞ 上的互质分解”.

模型匹配问题(model matching problem) 一种有理函数矩阵的逼近问题.已知真实有理矩阵 $T_i(s)\in RH_\infty$,$i=1,2,3$,求一个矩阵 $Q(s)\in RH_\infty$,使得

$$\|T_1(s)-T_2(s)Q(s)T_3(s)\|_\infty$$

达到极小,这就是典型的模型匹配问题.H_∞ 控制的标准问题可以化成一个模型匹配问题.

已知广义受控对象 $P(s)$,$P_{22}(s)$ 的左、右互质分解分别为 $(\tilde{M}_2(s),\tilde{N}_2(s))$ 和 $(M_2(s),N_2(s))$.定义:

$$T_1(s)\triangleq P_{11}(s)+P_{12}(s)M_2(s)\tilde{Y}_2(s)P_{21}(s);$$
$$T_2(s)\triangleq P_{12}(s)M_2(s);$$
$$T_3(s)\triangleq\tilde{M}_2(s)P_{21}(s);$$

则矩阵 $T_i(s)\in RH_\infty, i=1,2,3$. 如果取镇定控制器

$$K(s)=(Y_2(s)-M_2(s)Q(s))(X(s)-N_2(s)Q(s))^{-1}=(\tilde{X}_2(s)-Q(s)\tilde{N}_2(s))^{-1}(\tilde{Y}_2(s)-Q(s)\tilde{M}_2(s)),Q(s)\in RH_\infty\text{任意},$$

那么闭环系统从 $w(s)$到 $z(s)$之间的传递函数矩阵可表示成

$$\mathscr{F}_1(P(s),K(s))=T_1(s)-T_2(s)Q(s)T_3(s).$$

因此,可以说一个 H_∞控制的标准问题可以转化成一个模型匹配问题. 次优问题是指求 $Q(s)\in RH_\infty$,使得

$$\| T_1(s)-T_2(s)Q(s)T_3(s)\|_\infty<\gamma,$$

γ为事先给定的正数.

汉克尔算子(Hankel operator) 一个从 H_2 空间到它的正交补空间 $H_2^\perp$ 上的映射. 它的定义为:已知 $F(s)\in L_\infty$,带有符号 F 的汉克尔算子满足

$$\Gamma_F g=\Pi F g,$$

其中 $g\in H_2$ 是任意的,Π 表示 L_∞ 中的元到 $H_2^\perp$ 上的投影算子. 若 $F(s)\in RL_\infty$,那么汉克尔算子 Γ_F 有有限秩. 若 $F(s)\in RL_\infty$,那么 Γ_F 的范数可按如下步骤计算:

1. 求 $F(s)$的反稳定部分的最小实现

$$\left(\begin{array}{c|c}A & B\\ \hline C & 0\end{array}\right).$$

2. 解李亚普诺夫方程

$$AL_c+L_cA^{\mathrm{T}}=BB^{\mathrm{T}},$$
$$AL_0+L_0A=C^{\mathrm{T}}C.$$

3. 求矩阵 L_cL_0 的特征值.

4. $\|\Gamma_F\|=\lambda_{\max}(L_cL_0)$,其中 $\lambda_{\max}(\cdot)$表示矩阵的最大特征值.

奈哈里定理(Nehari theorem) 求从矩阵 $R(s)\in L_\infty$到 H_∞空间的距离的一个重要结果. 这个定理为:对于给定的矩阵 $R(s)\in L_\infty$,总存在矩阵 $X_0(s)\in H_\infty$,使得

$$\inf\{\|R(s)-X(s)\|_\infty;X(s)\in H_\infty\}=\|R(s)-X_0(s)\|_\infty,$$

并且 $\|R(s)-X_0(s)\|_\infty=\|\Gamma_R\|$.

这个定理说明一个不稳定的真实有理矩阵按照 L_∞范数可以用一个稳定的有理矩阵去近似,其近似误差的 L_∞范数正好等于该矩阵的汉克尔算子范数. 它可用来解模型匹配问题,因此可以把奈哈里定理看成是 H_∞控制问题求解的一个重要理论工具.

真有理矩阵的内外分解(inner-outer factorization for properreal-rational matrix) 有理函数矩阵的一类特殊的分解形式,它是根据 H_∞控制理论的需要而提出的. 一个真实有理函数矩阵 $G(s)\in RH_\infty$称为是内矩阵,如果 $G(s)^\sim G(s)=I$,这里 $G(s)^\sim=G^{\mathrm{T}}(-s)$. 显然,内矩阵是列满秩的. 称$G(s)\in RH_\infty$是外矩阵,如果对每个 $\mathrm{Re}(s)>0$ 的 s,$G(s)$都是行满秩的. 对外矩阵总存在一个在右半开复平面内解析的右逆. 内矩阵的列数小于或等于行数,外矩阵的列数大于或等于行数. 假设

$$G(s)=\left(\begin{array}{c|c}A & B\\ \hline C & D\end{array}\right),$$

那么 $G(s)$为内矩阵的充分必要条件是:

1. $B^{\mathrm{T}}X+D^{\mathrm{T}}C=0$.

2. $D^{\mathrm{T}}D=I$.

其中矩阵 X 是李亚普诺夫方程

$$A^{\mathrm{T}}X+XA+C^{\mathrm{T}}C=0$$

的解. 令 $G(s)\in RH_\infty$,如果存在一个内矩阵 $G_i(s)$和一个外矩阵 $G_o(s)$,使得

$$G(s)=G_i(s)G_o(s),$$

则称这个因式分解为 $G(s)$的内外分解. 事实上,在 RH_∞中的每个矩阵都有内外分解. 称矩阵$G(s)\in RH_\infty$是协内的或协外的,如果 $G^{\mathrm{T}}(s)$是内矩阵或外矩阵. 任意 $G(s)\in RH_\infty$都有协内外分解,即存在一个协外矩阵 $G_{co}(s)$和一个协内矩阵 $G_{ci}(s)$,使得

$$G(s)=G_{co}(s)G_{ci}(s).$$

由 $G^{\mathrm{T}}(s)$的内外分解可以得到 $G(s)$的协内外分解.

模型匹配问题的分类(classification of model-matching problem) 根据 $T_i(s)(i=1,2,3)$不同秩情况的模型匹配问题所对应的不同类别. 由于一个 H_∞控制问题可以化成模型匹配问题,因此求解模型匹配问题就成了解 H_∞控制的标准问题的重要途径. 为求解模型匹配问题,首先对它进行分类,然后根据不同的类型采用不同的处理方法. 这里将对模型匹配问题分成三类来讨论. 为此令 $T_1(s)$,$T_2(s)$和 $T_3(s)$分别为 RH_∞中的 $p\times l$,$p\times k$ 和 $h\times l$ 阶的矩阵.

1. 一块优化问题. 假设 $T_2(s)$行满秩,$T_3(s)$列满秩,显然有 $p\leqslant k, h\geqslant l$. 则存在 $H_2(s),H_2^{-1}(s)\in RH_\infty,N_3(s),N_3^{-1}(s)\in RH_\infty$,使得

$$T_2(s)=[T_{21}(s),0]H_2(s),$$
$$T_3(s)=N_3(s)\begin{pmatrix}T_{31}(s)\\0\end{pmatrix},$$

其中 $T_{21}(s)\in RH_\infty$为 $p\times p$ 阶矩阵,$T_{31}(s)\in RH_\infty$为 $l\times l$ 阶矩阵,均满秩. 于是

$$T_1(s)-T_2(s)Q(s)T_3(s)=T_1(s)-[T_{21}(s),0]H_2(s)Q(s)N_3(s)\begin{pmatrix}T_{31}(s)\\0\end{pmatrix}=T_1(s)-T_{21}(s)Q_1(s)T_{31}(s),$$

其中 $Q_1(s)=H_2(s)Q(s)N_3(s)$. 显然,$Q_1(s)$和$Q_2(s)$有一一对应的关系. 对 $T_{21}(s)$和 $T_{31}(s)$分别进行内外分解和协内外分解,得

$$T_{21}(s)=T_{21}^{i}(s)T_{21}^{o}(s),$$
$$T_{31}(s)=T_{31}^{co}(s)T_{31}^{ci}(s),$$

其中$(T_{21}^{o}(s))^{-1}\in RH_\infty$，$(T_{31}^{co}(s))^{-1}\in RH_\infty$. 于是有

$$\begin{aligned}&\|T_1(s)-T_2(s)Q(s)T_3(s)\|_\infty\\=&\|T_1(s)-T_{21}^{i}(s)Q_1(s)T_{31}^{ci}(s)\|_\infty\\=&\|R_1(s)-X_1(s)\|_\infty,\end{aligned}$$

其中

$$R_1(s)=(T_{21}^{i}(s))^{-1}T_1(s)(T_{31}^{ci}(s))^{-1}\in RL_\infty,$$
$$X_1(s)=T_{21}^{o}(s)H_2(s)Q(s)N_3(s)T_{31}^{co}(s)\in RH_\infty.$$

这样，模型匹配问题就被转化成这样一个问题，即已知$R_1(s)\in RL_\infty$，求$X_1(s)\in RH_\infty$，使得

$$\|R_1(s)-X_1(s)\|_\infty$$

达到极小，或使

$$\|R_1(s)-X_1(s)\|_\infty<\gamma,$$

这里γ是事先给定的正数. 前者为一块最优化问题，后者为一块次优问题.

2. 二块优化问题. 假设$T_2(s)$列满秩，$T_3(s)$为满秩方阵. 对$T_2(s)$，$T_3(s)$分别做内外分解和协内外分解：

$$T_2(s)=T_2^{i}(s)T_2^{o}(s),$$
$$T_3(s)=T_3^{co}(s)T_3^{ci}(s),$$

其中$T_2^{o}(s)$，$(T_2^{o}(s))^{-1}\in RH_\infty$，$T_3^{co}(s)$，$(T_3^{co}(s))^{-1}\in RH_\infty$. 可以证明，存在矩阵$T_2^{\perp}(s)\in RH_\infty$，使得$[T_2^{i}(s),T_2^{\perp}(s)]$为一方的内矩阵，于是有

$$\begin{aligned}&\|T_1(s)-T_2(s)Q(s)T_3(s)\|_\infty\\=&\|T_1(s)-T_2^{i}(s)T_2^{o}(s)Q(s)T_3^{co}(s)T_3^{ci}(s)\|_\infty\\=&\|T_1(s)-T_2^{i}(s)Q_1(s)T_3^{ci}(s)\|_\infty\\=&\left\|T_1(s)-[T_2^{i}(s),T_2^{\perp}(s)]\begin{pmatrix}Q_1(s)\\0\end{pmatrix}T_3^{ci}(s)\right\|_\infty\\=&\left\|\begin{pmatrix}R_1(s)-X_1(s)\\R_2(s)\end{pmatrix}\right\|_\infty,\end{aligned}$$

其中

$$R_1(s)=(T_2^{i}(s))^{\sim}T_1(s)(T_3^{ci}(s))^{\sim}\in RL_\infty,$$
$$R_2(s)=(T_2^{\perp}(s))^{\sim}T_1(s)(T_3^{ci}(s))^{\sim}\in RL_\infty,$$
$$X_1(s)=Q_1(s)=T_2^{o}(s)Q(s)T_3^{co}(s)\in RH_\infty.$$

显然，$X_1(s)$与$Q(s)$是一一对应的. 因此在这种情况下，模型匹配问题就转化成了两块问题，即已知矩阵$R_1(s)$，$R_2(s)\in RL_\infty$，求$X_1(s)\in RH_\infty$，使得

$$\left\|\begin{pmatrix}R_1(s)-X_1(s)\\R_2(s)\end{pmatrix}\right\|_\infty$$

达到极小或小于事先给定的正数γ，前者为二块最优化问题，后者为二块次优问题. 同样，当$p=k$，$h<l$时，即$T_2(s)$为满秩方阵，$T_3(s)$行满秩时，可得到另一个两块优化问题，即已知矩阵$R_1(s)$，$R_3(s)\in RL_\infty$，求$X_1(s)\in RH_\infty$，使得

$$\|[R_1(s)-X_1(s),\ R_3(s)]\|_\infty$$

达到极小或小于事先给定的正数γ.

3. 四块优化问题. 假设$T_2(s)$列满秩，$T_3(s)$行满秩. 对$T_2(s)$和$T_3(s)$分别进行内外分解和协内外分解，有

$$T_2(s)=T_2^{i}(s)T_2^{o}(s),$$
$$T_3(s)=T_3^{co}(s)T_3^{ci}(s),$$

其中$T_2^{o}(s)$，$(T_2^{o}(s))^{-1}\in RH_\infty$，$T_3^{co}(s)$，$(T_3^{co}(s))^{-1}\in RH_\infty$. 取$T_2^{\perp}(s)\in RH_\infty$，使$[T_2^{i}(s),T_2^{\perp}(s)]$为方内矩阵，取$T_3^{\perp}(s)\in RH_\infty$，使

$$\begin{pmatrix}T_3^{ci}(s)\\T_3^{\perp}(s)\end{pmatrix}$$

为方协内矩阵. 于是有

$$\begin{aligned}&\|T_1(s)-T_2(s)Q(s)T_3(s)\|_\infty\\=&\|T_1(s)-T_2^{i}(s)T_2^{o}(s)Q(s)T_3^{co}(s)T_3^{ci}(s)\|_\infty\\=&\left\|T_1(s)-[T_2^{i}(s),T_2^{\perp}(s)]\cdot\begin{pmatrix}X_1(s)&0\\0&0\end{pmatrix}\begin{pmatrix}T_3^{ci}(s)\\T_3^{\perp}(s)\end{pmatrix}\right\|_\infty\\=&\left\|\begin{pmatrix}R_1(s)-X_1(s)&R_3(s)\\R_2(s)&R_4(s)\end{pmatrix}\right\|_\infty,\end{aligned}$$

其中

$$\begin{pmatrix}R_1(s)&R_3(s)\\R_2(s)&R_4(s)\end{pmatrix}=\begin{pmatrix}(T_2^{i}(s))^{-1}\\(T_2^{\perp}(s))^{-1}\end{pmatrix}T_1(s)[(T_3^{ci}(s))^{-1},(T_3^{\perp}(s))^{-1}],$$
$$X_1(s)=T_2^{o}(s)Q(s)T_3^{co}(s).$$

因此，在这种情况下，模型匹配问题就转化成四块优化问题，即已知$R_i(s)\in RL_\infty$，$i=1,2,3,4$，求$X_1(s)\in RH_\infty$，使得

$$\left\|\begin{pmatrix}R_1(s)-X_1(s)&R_3(s)\\R_2(s)&R_4(s)\end{pmatrix}\right\|_\infty$$

达到极小或小于事先给定的正数γ. 前者为最优化问题，后者为次优问题. 综上所述可以看出，模型匹配问题主要是研究一块、二块或四块优化问题的解. 到目前为止，已有一些求解办法，但都很繁杂. 从系统设计的角度来说，解H_∞控制问题的一条重要途径就是研究各种求解模型匹配问题的方法.

系统不确定性的频域表示(frequency domain representation of system uncertainty) 利用频域语言(或传递函数矩阵)来表示系统的各种不确定性. 由于建模等各种因素都会使系统产生不确定性. 对于定常线性系统而言，按频域方法描述系统的不确定性一般有三种类型：

1. 加性不确定性. 如果用$G_\Delta(s)$表示带有不确定性的受控对象的传递函数矩阵，$\Delta_A(s)$表示系统的加性摄动，那么带有加性不确定性的摄动系统可表示成

$$G_\Delta(s)=G(s)+\Delta_A(s),$$

其中 $G(s)$ 为标称系统的传递函数矩阵.

2. 乘性不确定性. 如果用 $\Delta_P(s)$ 表示系统的乘性摄动，那么带有乘性不确定性的摄动系统可表示成 $G_\Delta(s)=(I+\Delta_P(s))G(s)$.

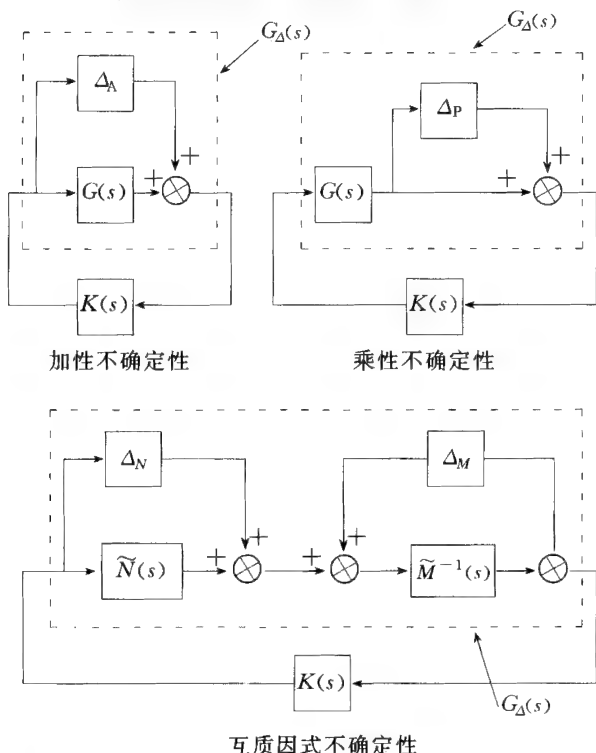
3. 互质因式不确定性. 假设标称系统 $G(s)$ 的左互质分解为 $\widetilde{M}^{-1}(s)\widetilde{N}(s)$，$\Delta_M(s)$ 和 $\Delta_N(s)$ 分别表示互质因式的不确定性，

$$\widetilde{N}_\Delta(s)=\widetilde{N}(s)+\Delta_N(s),$$
$$\widetilde{M}_\Delta(s)=\widetilde{M}(s)+\Delta_M(s),$$

那么带有互质因式不确定性的摄动系统可表示成

$$G_\Delta(s)=\widetilde{M}_\Delta^{-1}(s)\widetilde{N}_\Delta(s)$$
$$=(\widetilde{M}(s)+\Delta_M(s))^{-1}(\widetilde{N}(s)+\Delta_N(s)).$$

带有摄动的系统方块图如下图所示.



带有不确定性的受控对象的传递函数矩阵可以用上面线性分式变换表示，即

$$G_\Delta(s)=\mathscr{F}_u(P(s),\Delta(s)).$$

对于加性不确定性，有

$$P(s)=\begin{pmatrix}0 & I\\ I & G(s)\end{pmatrix},\ \Delta(s)=\Delta_A(s);$$

对于乘性不确定性，有

$$P(s)=\begin{pmatrix}0 & G(s)\\ I & G(s)\end{pmatrix},\ \Delta(s)=\Delta_P(s);$$

对于互质因式不确定性，有

$$P(s)=\begin{pmatrix}0 & I\\ \widetilde{M}^{-1}(s) & G(s)\\ \widetilde{M}^{-1}(s) & G(s)\end{pmatrix},$$
$$\Delta(s)=[\Delta_N(s),-\Delta_M(s)].$$

显然，标称系统为 $G(s)=\mathscr{F}_u(P(s),0)$.

鲁棒镇定定理(robust stabilization theorem) 关于带有不确定性的受控对象存在鲁棒镇定控制器的条件的一个定理. 对一个带有不确定性的受控对象，如果存在一个控制器，使得闭环系统总是内部稳定的，那么就称这个控制器为该系统的鲁棒镇定控制器，又称这个控制器鲁棒镇定这个不确定系统. 假设摄动 $\Delta(s)\in\mathscr{D}_\varepsilon=\mathscr{D}_{s_\varepsilon}\cup\mathscr{D}_{u_\varepsilon}$，其中

$$\mathscr{D}_{s_\varepsilon}\triangleq\{\Delta(s):\Delta(s)\in RH_\infty;\ \|\Delta(s)\|_\infty<\varepsilon\},$$
$$\mathscr{D}_{u_\varepsilon}\triangleq\{\Delta(s):\Delta(s)\in RL_\infty;\eta(\mathscr{F}_u(P(s),0))$$
$$=\eta(\mathscr{F}_u(P(s),\Delta(s)));\ \|\Delta(s)\|_\infty<\varepsilon\},$$

这里 $\eta(\cdot)$ 表示一个传递函数矩阵在右半闭复平面内的极点个数，$\eta(\mathscr{F}_u(P(s),0))=\eta(\mathscr{F}_u(P(s),\Delta(s)))$ 表示摄动前后系统在右半闭复平面内的极点个数不变，$P(s)$ 是给定的标称系统的传递函数矩阵.

鲁棒镇定定理：假设广义对象 $P(s)$ 没有不稳定的零极相消，则对一切摄动 $\Delta(s)\in\mathscr{D}_\varepsilon$，控制器 $K(s)$ 镇定不确定系统 $\mathscr{F}_u(P(s),\Delta(s))$（参见“线性分式变换”）的充分必要条件是：

1. $K(s)$ 镇定标称系统 $\mathscr{F}_u(P(s),0)$.

2. $\|\mathscr{F}_l(P(s),K(s))\|_\infty\leqslant 1/\varepsilon$.

关于鲁棒镇定定理的几点说明：

1. 令 $P(s)$ 的状态空间实现为

$$P(s)=\left(\begin{array}{c|cc}A & B_1 & B_2\\ \hline C_1 & D_{11} & D_{12}\\ C_2 & D_{21} & D_{22}\end{array}\right),$$

那么所谓广义对象 $P(s)$ 没有不稳定的零极相消是指 (A,B_2) 能稳，(C_2,A) 能检测. 它是用来保证 $P(s)$ 的镇定控制器存在的必不可少的条件.

2. 定理的第一个结论是说，控制器 $K(s)$ 鲁棒镇定系统 $\mathscr{F}_u(P(s),\Delta(s))$，首先应能镇定标称系统 $\mathscr{F}_u(P(s),0)$. 第二个结论只是说在没有摄动的情况下，闭环系统传递函数矩阵的 H_∞ 范数所满足的条件. 从小增益定理的观点来看，这个条件是显然的. 因此有人也称该定理为鲁棒小增益定理. 但这里是一个充分必要条件，而原来的小增益定理只是一个充分条件.

3. 几种特殊情况下的推论. 对于加性不确定性摄动，定理可改叙为：对所有 $\Delta_A(s)\in\mathscr{D}_\varepsilon$，控制器 $K(s)$ 鲁棒镇定系统 $G_\Delta(s)=G(s)+\Delta_A(s)$ 的充分必要条件是：$K(s)$ 镇定 $G(s)$，且

$$\|K(s)(I-G(s)K(s))^{-1}\|_\infty\leqslant 1/\varepsilon.$$

对于乘性不确定性摄动，定理可改叙为：对所有 $\Delta_P(s)\in\mathscr{D}_\varepsilon$，控制器 $K(s)$ 鲁棒镇定系统

$$G_\Delta(s)=(I+\Delta_P(s))G(s)$$

的充分必要条件是：$K(s)$ 镇定 $G(s)$，且

$$\|G(s)K(s)(I-G(s)K(s))^{-1}\|_\infty\leqslant 1/\varepsilon.$$

对于互质因式不确定性摄动，定理可改叙为：对所有$[\Delta_N(s),\Delta_M(s)]\in\mathscr{D}_\varepsilon$，控制器$K(s)$鲁棒镇定系统$G_\Delta(s)=(\tilde{M}(s)+\Delta_M(s))^{-1}(\tilde{N}(s)+\Delta_N(s))$的充分必要条件是：$K(s)$镇定$G(s)$；且

$$\left\|\begin{pmatrix}K(s)(I-G(s)K(s))^{-1}\tilde{M}^{-1}(s)\\(I-G(s)K(s))^{-1}\tilde{M}^{-1}(s)\end{pmatrix}\right\|_\infty\leqslant 1/\varepsilon.$$

4. 从定理可以看出，对于标称系统

$$G(s)=\mathscr{F}_u(P(s),0)$$

来说，任意给定的一个镇定控制器$K(s)$都可以确定一个允许的摄动范围

$$\varepsilon'=(\|\mathscr{F}_l(P(s),K(s))\|_\infty)^{-1},$$

即对任意$\Delta(s)\in\mathscr{D}_{\varepsilon'}$，$K(s)$鲁棒镇定摄动系统$\mathscr{F}_u(P(s),\Delta(s))$. 如果让$K(s)$取遍镇定标称系统

$$G(s)=\mathscr{F}_u(P(s),0)$$

的控制器，那么

$$\varepsilon_{\max}=(\inf_{K(s)}\|\mathscr{F}_l(P(s),K(s))\|_\infty)^{-1}$$

就决定了标称系统$G(s)=\mathscr{F}_u(P(s),0)$允许的最大摄动范围. 因此，鲁棒镇定问题可以转化为H_∞最优控制问题.

有界实引理(bounded real lemma)　描述有界实系统状态空间实现特定性质的一个引理. 它在讨论H_∞控制的次优解中起着关键性作用. 已知定常线性系统

$$G(s)=\left(\begin{array}{c|c}A&B\\\hline C&D\end{array}\right),$$

其中A,B,C,D分别是$n\times n,n\times r,m\times n$和$m\times r$阶常值矩阵. 定义$2n\times 2n$阶哈密顿矩阵

$$H=\begin{pmatrix}A+BR^{-1}D^{\mathrm{T}}C & BR^{-1}B^{\mathrm{T}}\\-C^{\mathrm{T}}C-C^{\mathrm{T}}DR^{-1}D^{\mathrm{T}}C & -A^{\mathrm{T}}-C^{\mathrm{T}}DR^{-1}B^{\mathrm{T}}\end{pmatrix},$$

其中$R=\gamma^2I-D^{\mathrm{T}}D$，$\gamma$为给定的正常数.

有界实引理：在A渐近稳定的前提下，下面的叙述是等价的：

1. $\|G(s)\|_\infty<\gamma$.

2. $\|D\|<\gamma$，并且矩阵H在虚轴上没有特征值.

3. $\|D\|<\gamma$，且存在$P=P^{\mathrm{T}}\geqslant 0$满足黎卡提矩阵方程

$$P(A+BR^{-1}D^{\mathrm{T}}C)+(A+BR^{-1}D^{\mathrm{T}}C)^{\mathrm{T}}P+PBR^{-1}B^{\mathrm{T}}P+C^{\mathrm{T}}(I+DR^{-1}D^{\mathrm{T}})C=0,$$

并且$A+BR^{-1}(D^{\mathrm{T}}C+B^{\mathrm{T}}P)$是渐近稳定的.

H_2控制问题(H_2 control problem)　按H_2范数指标设计控制器的问题. 对于给定的广义受控对象$P(s)$，求容许控制器$K(s)$，使得闭环系统内部稳定且传递函数矩阵$\mathscr{F}_l(P(s),K(s))$(参见"线性分式变换")的H_2范数达到极小. 假设$P(s)$的状态空间实现为

$$P(s)=\left(\begin{array}{c|cc}A&B_1&B_2\\\hline C_1&D_{11}&D_{12}\\C_2&D_{21}&D_{22}\end{array}\right),$$

不失一般性，可设$D_{11}=0,D_{22}=0$，且

1. (A,B_1)能稳，(C_1,A)能检测.

2. (A,B_2)能稳，(C_2,A)能检测.

3. $D_{12}^{\mathrm{T}}[C_1,D_{12}]=[0,I]$.

4. $\begin{pmatrix}B_1\\D_{21}\end{pmatrix}D_{21}^{\mathrm{T}}=\begin{pmatrix}0\\I\end{pmatrix}$.

对于给定的广义受控对象$P(s)$，假设它满足上述条件1～4，则存在惟一的容许控制器$K(s)$，使得闭环系统内部稳定，且$\|\mathscr{F}_l(P(s),K(s))\|_2$达到极小. 这时

$$K(s)\triangleq\left(\begin{array}{c|c}\hat{A}_2&-L_2\\\hline F_2&0\end{array}\right),$$

其中

$$\hat{A}_2\triangleq A+B_2F_2+L_2C_2,$$
$$F_2\triangleq -B_2^{\mathrm{T}}X_2,$$
$$L_2\triangleq -Y_2C_2^{\mathrm{T}}.$$

这里X_2和Y_2分别是如下代数黎卡提矩阵方程的对称非负定解：

$$A^{\mathrm{T}}X_2+X_2A-X_2B_2B_2^{\mathrm{T}}X_2+C_1^{\mathrm{T}}C_1=\theta,$$
$$AY_2+Y_2A^{\mathrm{T}}-Y_2C_2^{\mathrm{T}}C_2Y_2+B_1B_1^{\mathrm{T}}=\theta,$$

并且

$$\begin{aligned}&\|\mathscr{F}_l(P(s),K(s))\|_2^2\\=&\|P_{\mathrm{C}}(s)B_1\|_2^2+\|F_2P_{\mathrm{f}}(s)\|_2^2\\=&\|P_{\mathrm{C}}(s)L_2\|_2^2+\|C_1P_{\mathrm{f}}(s)\|_2^2\end{aligned}$$

为闭环系统传递函数矩阵$\mathscr{F}_l(P(s),K(s))$的最小H_2范数，其中

$$P_{\mathrm{C}}(s)\triangleq\left(\begin{array}{c|c}A_{F_2}&I\\\hline C_{1F_2}&0\end{array}\right),\quad P_{\mathrm{f}}(s)\triangleq\left(\begin{array}{c|c}A_{L_2}&B_{1L_2}\\\hline I&0\end{array}\right),$$

$$A_{F_2}\triangleq A+B_2F_2,\quad C_{1F_2}\triangleq C_1+D_{12}F_2,$$
$$A_{L_2}\triangleq A+L_2C_2,\quad B_{1L_2}\triangleq B_1+L_2D_{21},$$

且A_{F_2}和A_{L_2}都是渐近稳定的.

H_2控制问题的次优解(suboptimal solution to H_2 control problem)　H_2控制问题的一种近似最优解. 对于给定的广义受控对象的$P(s)$，求使得闭环系统的传递函数矩阵$\mathscr{F}_l(P(s),K(s))$(参见"线性分式变换")的H_2范数小于事先给定的正常数γ的容许控制器$K(s)$. 这里所谓容许控制器是指使闭环系统内部稳定的控制器. 对$P(s)$的假设同前，则有如下关于H_2控制问题次优解的一个定理.

使得$\|\mathscr{F}_l(P(s),K(s))\|_2<\gamma$的所有容许控制器$K(s)$为

$$u(s)=\mathscr{F}_l(M_2(s),Q(s))y(s),$$

即$K(s)=\mathscr{F}_l(M_2(s),Q(s))$，其中

$$M_2(s) \triangleq \left(\begin{array}{c|cc} \hat{A}_2 & -L_2 & B_2 \\ \hline F_2 & 0 & I \\ -C_2 & I & 0 \end{array}\right),$$

$$Q(s) \in RH_2,$$

且

$\| Q(s) \|_2^2 < \gamma^2 I - (\| P_C(s) B_1 \|_2^2 + \| F_2 P_f(s) \|_2^2)$，这里 $\hat{A}_2, L_2, F_2, P_C(s)$ 和 $P_f(s)$ 的定义详见“H_2 控制问题”.

H_∞控制问题的次优解(suboptimal solution to H_∞ control problem)　H_∞控制问题的一种近似最优解. 对于给定的广义受控对象$P(s)$，求所有的容许控制器 $K(s)$，使得闭环系统的传递函数矩阵 $\mathscr{F}_l(P(s),K(s))$（参见“线性分式变换”）的 H_∞范数小于事先给定的正常数.

对 $P(s)$的假设同“H_2 控制问题”，则有如下关于 H_∞控制问题次优解的一个定理.

定理：对于给定的广义受控对象 $P(s)$有：

1. 存在一个容许控制器 $K(s)$，使得

$$\| \mathscr{F}_l(P(s),K(s)) \|_\infty < \gamma$$

的充分必要条件是下列条件成立：

1）代数黎卡提方程

$$A^T X + XA + X(\gamma^{-2} B_1 B_1^T - B_2 B_2^T) X + C_1^T C_1 = 0$$

有对称非负定解 X.

2）代数黎卡提方程

$$AY + YA^T + Y(\gamma^{-2} C_1^T C_1 - C_2^T C_2) Y + B_1 B_1^T = 0$$

有对称非负定解 Y.

3）$\rho(XY) < \gamma^2$，其中 $\rho(\cdot)$表示矩阵的谱半径.

当这些条件成立时，这样一个控制器为

$$K_2(s) = \left(\begin{array}{c|c} \hat{A}_\infty & -Z_\infty L_\infty \\ \hline F_\infty & 0 \end{array}\right),$$

这里

$$\hat{A}_\infty \triangleq A + \gamma^{-2} B_1 B_1^T X + B_2 F_\infty + Z_\infty L_\infty C_2,$$

$$F_\infty \triangleq -B_2^T X, L_\infty \triangleq -Y C_2^T,$$

$$Z_\infty \triangleq (I - \gamma^{-2} YX)^{-1}.$$

2. 如果上述条件 1)和 2)都被满足，则所有使得

$$\| \mathscr{F}_l(P(s),K(s)) \|_\infty < \gamma$$

的容许控制器 $K(s)$可表示成

$$K(s) = \mathscr{F}_l(M_\infty(s), Q(s)),$$

其中

$$M_\infty(s) \triangleq \left(\begin{array}{c|cc} \hat{A}_\infty & -Z_\infty L_\infty & Z_\infty B_2 \\ \hline F_\infty & 0 & I \\ -C_2 & I & 0 \end{array}\right),$$

$$Q(s) \in RH_\infty \text{ 且 } \| Q(s) \|_\infty < \gamma.$$

线性矩阵不等式(linear matrix inequality)　简称 LMI. 未知向量满足的一个矩阵约束关系. 是指未知向量满足形如

$$F(x) = F_0 + x_1 F_1 + \cdots + x_N F_N < 0 (\leqslant 0)$$

的一个不等式，其中 $x = (x_1, x_2, \cdots, x_N)$为一未知的向量，$F_0, F_1, \cdots, F_N$ 为已知的对称矩阵，“$F(x) < 0$”表示 $F(x)$为负定矩阵. 线性矩阵不等式 $F(x) < 0$ 是关于 x 的一个凸约束，它的解集（称为可解集）为 $\mathbb{R}^N$ 中的一个凸子集，寻找 $F(x) < 0$ 解的问题（如果存在）是一凸最优化问题. 在大部分控制问题中，LMI 常以

$$L(X_1, X_2, \cdots, X_n) < R(X_1, X_2, \cdots, X_n)$$

的形式出现，其中 $L(\cdot)$与 $R(\cdot)$为其结构矩阵变量 $X_1, X_2, \cdots, X_n$ 的仿射函数.

LMI 的三个基本问题是：

1. 可解问题，即寻找 LMI：$F(x) < 0$ 解的问题.

2. 线性目标最小化问题，即

$$\text{Minimize } C^T x,\ \text{约束：} F(x) < 0.$$

3. 广义特征值最小问题，即 Minimize λ，

约束：$A(x) < \lambda B(x), B(x) > 0, C(x) < 0$.

控制中的线性矩阵不等式(linear matrix inequality in control)　控制问题中常见的或将问题经转化而成的线性矩阵不等式. LMI 在控制理论与控制工程中主要具有以下几个方面的应用：鲁棒控制，包括鲁棒稳定性与性能分析、二次稳定、参数相依李亚普诺夫稳定性、μ 分析、波波夫分析以及线性定常系统的输入和输出性质，如不变椭圆、衰减率等；多模型/多目标状态反馈设计；输出反馈 H_∞控制器综合，包括具有极点配置约束的混合 H_2/H_∞综合问题；最优化 LQG 控制；回路成形（loop-shaping）设计；模型降阶以及控制器逼近问题；时变参数相依系统的鲁棒增益调整(gain-scheduled)控制器的设计；随机系统的控制；加权插值问题等，这些问题经处理均可化成各种 LMI 问题.

LMI 问题可以通过有效的凸优化算法（如椭圆算法、内点算法）从数值上进行求解，MATLAB 包含有 LMI 控制工具箱，从而使得与 LMI 有关的一些控制问题能够得到处理和解决.

控制中 LMI 问题的难点在于在 LMI 约束的同时，可能还存在非凸的矩阵秩约束.

积分二次约束(integral quadratic constraint)　系统输入输出之间所受的一种二次形式的积分约束(IQC). 其时域形式最早见于超稳定性的理论中，频域形式由雅库波维奇（Якубович, Б. А.）在研究非线性系统时提出，在无源性分析中也起着重要的作用.

记 RL_∞为实系数真有理函数集，其中所有极点仅位于左半开复平面内的函数集，记为 RH_∞. 由 RL_∞（或 RH_∞）中的元素组成的 $m \times n$ 阶矩阵的集合记为

$$RL_\infty^{m \times n} (\text{或 } RH_\infty^{m \times n}).$$

$L_2^l[0,\infty)$是从$[0,\infty)$到 R^l 的所有平方可积的函数

空间，而 $L_{2e}^{l}[0,\infty)$ 是其扩展空间，它的元素只需在任何有限区间上可积即可. $\hat{f}(\mathrm{j}\omega)$ 表示 $f\in L_2^l[0,\infty)$ 的傅里叶变换；从一个 $L_{2e}^{a}[0,\infty)$ 空间至另一个 $L_{2e}^{b}[0,\infty)$ 算子 F 的算子增益由其范数给出；$P_T(T>0)$ 表示截断算子；算子 F 被称为是因果的，若对所有 $T>0, P_TFP_T=P_TF$.

设 $\Pi:\mathrm{j}\omega\to C^{(l+m)\times(l+m)}$ 是一取厄米特值的有界可测矩阵函数，其中 $\omega\in\mathrm{R}$.

$$\Delta:L_{2e}^{l}[0,\infty)\to L_{2e}^{m}[0,\infty)$$

为有界算子，若对所有 $v\in L_2^l[,\infty)$ 有

$$\sigma(v,\Delta(v))\triangleq\int_{-\infty}^{\infty}\begin{bmatrix}\hat{v}(\mathrm{j}\omega)\\ \Delta(\hat{v})(\mathrm{j}\omega)\end{bmatrix}^{*}\Pi(\mathrm{j}\omega)\begin{bmatrix}\hat{v}(\mathrm{j}\omega)\\ \Delta(\hat{v})\end{bmatrix}\mathrm{d}\omega>0$$

成立，则称 Δ 满足由 Π 确定的积分二次约束(IQC)，并记 $\Delta\in\mathrm{IQC}(\Pi)$，矩阵函数 Π 称为 IQC 的乘子.

系统分析中的积分二次约束方法(system analysis via integral quadratic constraint)　基于积分二次约束对控制系统进行分析与综合的方法. 积分二次约束所刻画的系统是一个系统族，它既可刻画系统中的非线性环节，又可刻画近年来得到广泛研究的结构与非结构的不确定性，从而有可能在研究系统不确定时抓住同一性，并在统一的框架下加以研究，同时，借助正实引理，可将多数的积分二次约束的检验问题转化为线性矩阵不等式，通过新近发展的内点多项式算法可以有效地进行计算.

考虑反馈内联系统

$$v=Gu+f,\ u=\Delta(v)+e,$$

这里 $f\in L_{2e}^{l}[0,\infty)$, $e\in L_{2e}^{m}[0,\infty)$, G 与 Δ 分别是定义于 $L_{2e}^{m}[0,\infty)$ 和 $L_{2e}^{l}[0,\infty)$ 上的两个因果算子. 假设 G 是一个其传递函数矩阵在 $RH_{\infty}^{l\times m}$ 中的线性时不变算子，Δ 具有有界的增益. 若

1. 对每一 $r\in[0,1]$, G 与 $r\Delta$ 的内联是适定的.
2. 对每一 $r\in[0,1]$, $r\Delta\in\mathrm{IQC}(\Pi)$.
3. 存在 $\varepsilon>0$，使

$$\begin{bmatrix}G(\mathrm{j}\omega)\\ I\end{bmatrix}^{*}\Pi(\mathrm{j}\omega)\begin{bmatrix}G(\mathrm{j}\omega)\\ I\end{bmatrix}\leqslant-\varepsilon I,\ \forall\,\omega\in\mathrm{R},$$

则 G 与 Δ 的反馈内联是稳定的.

在上述结论中，若令

$$\Pi(\mathrm{j}\omega)=\begin{bmatrix}I & 0\\ 0 & -I\end{bmatrix},$$

则得到小增益定理的一种形式；当

$$\Pi(\mathrm{j}\omega)=\begin{pmatrix}\varepsilon & I\\ I & -\varepsilon/\|\Delta\|^2\end{pmatrix},$$

则得到无源性定理；当

$$\Pi(\mathrm{j}\omega)=\begin{bmatrix}0 & M^{*}\\ M & 0\end{bmatrix},$$

则得到输入-输出稳定. 与古典乘子定理相比较，乘子的分解性条件可以避免，为此而加强的条件仅需对每一 $r\in[0,1]$ 时 G 与 $r\Delta$ 的反馈内联是适定的. 这一条件对大多数应用来说是较弱的.

绝对稳定性

绝对稳定性(absolute stability)　具有非线性不确定性的控制系统的一种稳定性. 在绝对稳定性理论中，时常研究形如

$$\begin{cases}\dot{x}=Ax+Bu,\\ u=-\varphi(\sigma),\\ \sigma=C^{\mathrm{T}}x\end{cases}$$

的控制系统的稳定性，式中 $x\in\mathrm{R}^n$, $u,\sigma\in\mathrm{R}^m$, $A\in\mathrm{R}^{n\times n}$, $B,C\in\mathrm{R}^{n\times m}$, $\varphi(\sigma)$ 是 σ 的连续向量函数，$\varphi(0)=0$. 研究的基本问题是在对 $\varphi(\sigma)$ 了解甚少，只知道它属于某一函数类的情况下，如何从控制系统的线性部分参数确定闭环非线性系统的全局渐近稳定性. 常取的函数类之一是：$\mathscr{F}_{\mu}=\{\varphi|\varphi=(\varphi_1,\varphi_2,\cdots,\varphi_m)^{\mathrm{T}}$, $\varphi_i(\sigma)=\varphi_i(\sigma_i)$ 是 σ_i 的连续函数，$0<\sigma_i\varphi_i(\sigma)<\mu_i\sigma_i^2$, $i=1,2,\cdots,m\}$，式中 $\mu=(\mu_1,\mu_2,\cdots,\mu_m)^{\mathrm{T}}$. 如果系统对于 $\mathscr{F}_{\mu}$ 中每一个函数 $\varphi(\sigma)$ 是全局渐近稳定的，则称系统关于 $\mathscr{F}_{\mu}$ 绝对稳定.

绝对稳定性不同于李亚普诺夫意义下的稳定性. 李亚普诺夫意义下的稳定性研究的是单个确定系统的稳定性，而绝对稳定性研究的是一类具不确定非线性特性的系统的全局渐近稳定性. 绝对稳定性理论是由于分析和设计控制系统和其他实际系统的需要而产生和发展起来的. 1944 年，苏联学者鲁里叶(Лурье, А. И.)和波斯特尼考夫(Постников, В. Н.)从一类实际控制系统出发，首次提出和研究了绝对稳定性问题，开创了研究一类非线性控制系统稳定性的新方法. 1949 年，苏联学者阿依热尔曼(Айзерман, М. А.)提出了一个有关绝对稳定性判据的猜想. 虽然该猜想不成立，但它促进了对绝对稳定性的研究. 1961 年，罗马尼亚学者波波夫(Попов, В. М.)提出了绝对稳定性的频率域判据，开创了研究绝对稳定性的新途径，并使绝对稳定性理论更适合于工程实际应用. 随着实际应用的扩大，所研究的系统已由常微分方程描述的系统扩大到由差分方程和泛函微分方程等所描述的系统. 绝对稳定性已成为控制系统稳定性理论的重要组成部分.

鲁里叶问题(Lur'e problem)　绝对稳定性研究中的一个著名问题. 1944 年，苏联学者鲁里叶(Лурье, А. И.)和波斯特尼考夫(Постников, В. Н.)提出了一种研究非线性控制系统稳定性的新方法. 他们从许多控制系统出发，将系统的非线性部分孤立出来，使系统具有闭环控制的形式，如图 1 所示.

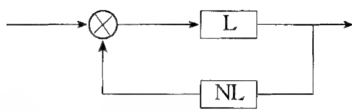


图1

图 1 中标有 L 的方框代表系统的线性部分，标有 NL 的方框代表系统的非线性部分. 鲁里叶和波斯特尼考夫提出了如下问题：在系统的非线性部分能够被孤立出来，不确知但满足一定的条件下，如何由系统的线性部分和非线性部分的几个界定参数来判别整个系统的全局稳定性？此问题称为鲁里叶问题. 假设图 1 中的非线性部分是一个单输入-单输出环节，如图 2，其输入 σ 和输出 $\varphi(\sigma)$ 都是标量. 设有函数类 $\mathscr{F}(\mu_1,\mu_2)=\{f|f$ 在R 上连续，$\mu_1\sigma^2<\sigma f(\sigma)<\mu_2\sigma^2,\sigma\neq 0\}$，其中 μ_1,μ_2 是给定的. 此时，鲁里叶问题就变为系统的线性部分应满足什么条件，才能保证对所有的 $\varphi\in\mathscr{F}(\mu_1,\mu_2)$ 系统是全局渐近稳定的？鲁里叶问题引起了对绝对稳定性的研究，促进了控制系统稳定性理论的发展.

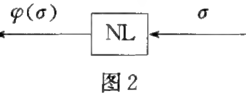


图 2

阿依热尔曼猜想(Aizerman conjecture) 绝对稳定性研究中的一个著名问题. 1949 年，苏联学者阿依热尔曼(Айзерман, M. A.)对控制系统

$$\begin{cases}\dot{x}=Ax+bu,\\ u=-\varphi(\sigma),\\ \sigma=c^{\mathrm{T}}x,\end{cases}$$

式中 $x\in\mathbf{R}^n, A\in\mathbf{R}^{n\times n}, b,c\in\mathbf{R}^n, u,\sigma\in\mathbf{R}$，提出了一个猜想：如果对每一个 $\varphi(\sigma)=\varepsilon\sigma, \varepsilon\in(\mu_1,\mu_2)$，该系统是渐近稳定的，那么该系统关于函数类 $\mathscr{F}(\mu_1,\mu_2)=\{f|f$ 在 R 上连续，$\mu_1\sigma^2<\sigma f(\sigma)<\mu_2\sigma^2,\sigma\neq 0\}$ 就是绝对稳定的. 已有反例表明阿依热尔曼猜想一般是不成立的，但它推动了对绝对稳定性的研究，同时促使人们去寻找使该猜想成立的条件.

迈耶-卡尔曼-雅库波维奇引理(Meyer-Kalman-Yacubovich lemma) 关于线性系统 $\Sigma(C,A,B)$ 的李亚普诺夫矩阵方程与其传递函数阵为正实矩阵时的几个定理的统称. 这些定理是研究绝对稳定性的主要工具，其中一个定理说：如果 $A\in\mathbf{R}^{n\times n}, b,c\in\mathbf{R}^n$，$A$ 的特征值实部均小于零，(A,b) 为能控对，常数 $r\geqslant 0$，那么矩阵方程

$$\begin{cases}PA+A^{\mathrm{T}}P=-qq^{\mathrm{T}},\\ Pb-c=\sqrt{r}\,q\end{cases}$$

有解 $P\geqslant 0$ 和 $q\in\mathbf{R}^n$ 的充分必要条件是：

$$r+2c^{\mathrm{T}}(sI-A)^{-1}b$$

为正实有理函数.

迈耶(Meyer, K. R.)、卡尔曼(Kalman, R. E.)和雅库波维奇(Якубович, Б. А.)对这些定理的建立做出了贡献，故称为迈耶-卡尔曼-雅库波维奇引理，也简称 MKY 引理. MKY 引理在稳定性的时域研究与频域研究间架起了桥梁.

鲁里叶方法(Lur'e approach) 亦称 S 方法. 研究绝对稳定性的主要方法之一. 该方法由鲁里叶(Лурье, А. И.)和波斯特尼考夫(Постников, В. Н.)首创，主要用于研究控制系统

$$\begin{cases}\dot{x}=Ax+bu,\\ u=-\varphi(\sigma),\\ \sigma=c^{\mathrm{T}}x\end{cases}$$

关于函数族 $\mathscr{F}_\mu$ 的绝对稳定性，式中 $x\in\mathbf{R}^n, A\in\mathbf{R}^{n\times n}, b,c\in\mathbf{R}^n, \mathscr{F}_\mu=\{f|f$ 在 R 上连续，$0<\sigma f(\sigma)<\mu\sigma^2,\sigma\neq 0\}$. 它的特点是：

1. 采用二次型加非线性积分项形式的 V 函数，即

$$V(x)=x^{\mathrm{T}}Px+\beta\int_0^\sigma\varphi(\sigma)\mathrm{d}\sigma,$$

式中 P 为待定的正定矩阵，β 为待定的常数.

2. 由 $V(x)$ 的正定性和 $\frac{\mathrm{d}V(x)}{\mathrm{d}t}$ 的负定性得出系统关于 $\mathscr{F}_\mu$ 的绝对稳定的充分条件.

波波夫判据(Popov criterion) 绝对稳定性的一个频域形式的判据. 设非线性系统由方程

$$\dot{x}=Ax-b\varphi(\sigma),\quad \sigma=c^{\mathrm{T}}x$$

描述，其中非线性函数 $\varphi(\sigma)$ 在 R 上连续，并且满足不等式 $0\leqslant\sigma\varphi(\sigma)\leqslant\mu\sigma^2$. 设 $W(s)=c^{\mathrm{T}}(sI-A)^{-1}b$，如果存在实数 q，使得对所有 $\omega\geqslant 0$，不等式

$$\mathrm{Re}\left\{(1+\mathrm{j}q\omega)W(\mathrm{j}\omega)+\frac{1}{\mu}\right\}>0$$

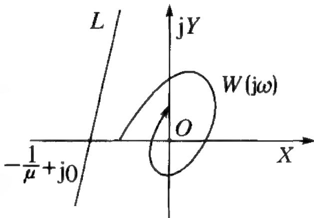
成立，则上述非线性系统绝对稳定. 波波夫判据可以表示成几何形式. 设 $\hat{W}(\mathrm{j}\omega)=X(\omega)+\mathrm{j}Y(\omega)$，式中 $X(\omega)=\mathrm{Re}\,(W(\mathrm{j}\omega)), Y(\omega)=\omega\,\mathrm{Im}(W(\mathrm{j}\omega))$. $\hat{W}(\mathrm{j}\omega)$ 称为修正频率响应或波波夫频率响应. 则波波夫判据可以表示为对所有 $\omega\geqslant 0$，

$$X(\omega)-qY(\omega)+\frac{1}{\mu}>0.$$

如图，在复平面上过点 $-1/\mu+\mathrm{j}0$，作斜率为 $1/q$ 的直线

$$L: x-qy+\frac{1}{\mu}=0$$

(称为波波夫直线)，则波波夫判据可以表示成：如果 $\hat{W}(\mathrm{j}\omega)$ 和原点在直线 L 的同一侧，则上述非线性系统绝对稳定.



波波夫判据是波波夫(Попов, В. М.)于 1961 年提出的.

圆判据(circle criterion) 绝对稳定性的一个

频域形式判据. 该判据利用复平面上的一个圆和系统线性部分的频率响应曲线的相互位置关系来判别系统的绝对稳定性，故有圆判据之称. 设非线性系统由方程

$$\dot{x} = Ax - b\varphi(\sigma),\ \sigma = c^{\mathrm{T}}x$$

描述，其中非线性函数在 R 上连续，且满足扇形条件 $\mu_1\sigma^2 \leqslant \sigma\varphi(\sigma) \leqslant \mu_2\sigma^2$.

设开环频率响应 $W(\mathrm{j}\omega) = c^{\mathrm{T}}(\mathrm{j}\omega - A)^{-1}b$. 用 $C(\mu_1, \mu_2)$ 表示复平面上以

$$-\frac{1}{2}\left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right) + \mathrm{j}0$$

为圆心，以

$$\frac{1}{2}\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right)$$

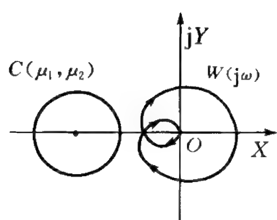


图 1

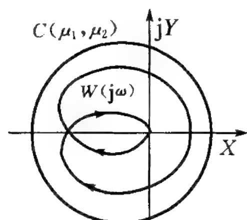


图 2

为半径的圆. 则圆判据称：如果下列条件之一成立，则上述非线性系统绝对稳定：

1. 设 $0 < \mu_1 < \mu_2$，频率响应曲线在圆 $C(\mu_1, \mu_2)$ 的外部. 若 A 有 r 个正实部特征值，则当 ω 从 $-\infty$ 增加到 $+\infty$ 时，$W(\mathrm{j}\omega)$ 逆时针方向绕圆 $C(\mu_1, \mu_2)$ 转 r 周. 在 A 有零实部特征值时，$W(\mathrm{j}\omega)$ 用一般奈奎斯特轨迹代替，如图 1.

2. 设 $\mu_1 < 0 < \mu_2$，A 的特征值实部均小于零，且 $W(\mathrm{j}\omega)$ 在圆 $C(\mu_1, \mu_2)$ 的内部，如图 2.

3. 设 $\mu_1 = 0 < \mu_2$，A 的特征值实部均小于零，且 $W(\mathrm{j}\omega)$ 位于直线 $C(0, \mu_2)$ 的右侧，其中直线 $C(0, \mu_2)$：$x = -1/\mu_2$，它由圆 $C(\mu_1, \mu_2)$ 当 μ_1 趋于零退化得到，如图 3.

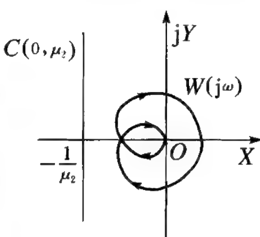


图 3

直接调节系统(direct control system) 一类非线性控制系统. 在绝对稳定性理论中，通常将控制系统

$$\begin{cases} \dot{x} = Ax + bu, \\ u = -\varphi(\sigma), \\ \sigma = c^{\mathrm{T}}x \end{cases}$$

称为直接调节系统，式中 $x, b, c \in \mathrm{R}^n$，$A \in \mathrm{R}^{n\times n}$，$u, \sigma \in \mathrm{R}$，$\varphi \in \mathscr{F}_\mu = \{f \mid f$ 在 R 上连续，$0 < \sigma f(\sigma) < \mu\sigma^2, \sigma \neq 0\}$. 这里“直接调节”的含义是：输出 $u = -\varphi(\sigma)$ 被直接用作线性部分的控制信号.

间接调节系统(indirect control system) 一类非线性控制系统. 在绝对稳定性理论中，除直接调节系统外，通称为间接调节系统. 间接调节系统有多种形式，例如，非线性控制系统

$$\begin{cases} \dot{x} = Dx + b\xi, \\ \dot{\xi} = u, \\ u = \varphi(\sigma), \\ \sigma = c^{\mathrm{T}}x - p\xi \end{cases}$$

称为间接调节系统，式中 $x, b, c \in \mathrm{R}^n$，$\xi, u, \sigma, p \in \mathrm{R}$，$D \in \mathrm{R}^{n\times n}$，$\varphi \in \mathscr{F}_\mu = \{f \mid f$ 在 R 上连续，$0 < \sigma f(\sigma) < \mu\sigma^2, \sigma \neq 0\}$. 这里“间接调节”的含义是：输出 $u = \varphi(\sigma)$ 经积分放大后才被用作线性部分的控制信号 ξ.

有理正实函数(rational positive real function) 一类复变量有理函数. 设 $f(s) = p(s)/q(s)$ 是复变量 s 的有理函数，$p(s)$ 和 $q(s)$ 为互质多项式. 如果 $f(s)$ 满足条件：

1. 当 s 为实数时，$f(s)$ 只要有定义就是实的；

2. 当 $\mathrm{Re}(s) \geqslant 0$ 时，只要 $f(s)$ 有定义就有

$$\mathrm{Re}(f(s)) \geqslant 0;$$

则称 $f(s)$ 为有理正实函数.

有理正实矩阵(rational positive real matrix) 一类复变量有理函数矩阵. 设 $G(s)$ 是复变量 s 的 $m \times m$ 有理函数矩阵. 如果 $G(s)$ 满足条件：

1. 当 $\mathrm{Re}(s) > 0$ 时，$G(s)$ 的所有元素都解析；

2. 对于实数 s，$G(s)$ 是实矩阵；

3. 当 $\mathrm{Re}(s) > 0$ 时，$G(s) + G^*(s) \geqslant 0$，式中 $G^*(s)$ 是 $G(s)$ 的共轭转置；

则称 $G(s)$ 为有理正实矩阵. 有理正实矩阵是有理正实函数概念的推广.

有理严格正实矩阵(rational strictly positive real matrix) 一类有理正实矩阵. 设 $G(s)$ 是复变量 s 的 $m \times m$ 实系数有理函数矩阵. 如果 $G(s)$ 满足条件：

1. 当 $\mathrm{Re}(s) \geqslant 0$ 时，$G(s)$ 的所有元素都解析；

2. 对所有实数 ω，$G(\mathrm{j}\omega) + G^*(\mathrm{j}\omega) > 0$；

3. $\lim\limits_{\omega\to\infty} \omega^2(G(\mathrm{j}\omega) + G^*(\mathrm{j}\omega)) > 0$；

则称 $G(s)$ 为有理严格正实矩阵.

有理弱严格正实矩阵(rational weakly strictly positive real matrix) 一类有理正实矩阵. 设 $G(s)$ 是复变量 s 的 $m \times m$ 实系数有理函数矩阵. 如果 $G(s)$ 满足条件：

1. 对 $\mathrm{Re}(s) \geqslant 0$，$G(s)$ 的所有元素都解析；

2. 对所有实数 ω，$G(\mathrm{j}\omega) + G^*(\mathrm{j}\omega) > 0$；

则称 $G(s)$ 为有理弱严格正实矩阵. 在有的文献里，将这里的有理弱严格正实矩阵称为有理严格正实矩

阵.

有理强严格正实矩阵(rational strongly strictly positive real matrix) 一类有理正实矩阵. 设 $G(s)$ 是复变量 s 的 $m\times m$ 实系数有理函数矩阵. 如果 $G(s)$ 满足条件:

1. 对 $\mathrm{Re}(s)\geqslant 0$, $G(s)$ 的所有元素都解析;
2. 对所有实数 ω, $G(\mathrm{j}\omega)+G^{*}(\mathrm{j}\omega)>0$;
3. $G(\infty)+G^{\mathrm{T}}(\infty)>0$;

则称 $G(s)$ 为有理强严格正实矩阵.

反馈正实系统(feedback positive real system) 一类线性控制系统. 用 $\Sigma(C,A,B)$ 表示线性控制系统

$$\begin{cases}\dot{x}=Ax+Bu,\\ y=C^{\mathrm{T}}x,\end{cases}$$

式中 $x\in\mathrm{R}^{n}$, $u,y\in\mathrm{R}^{m}$, $A\in\mathrm{R}^{n\times n}$, $B,C\in\mathrm{R}^{n\times m}$. 如果存在矩阵 $P>0$, $Q\geqslant 0$ 和 K, 使得 $A+BK$ 是稳定矩阵, 并且满足矩阵方阵

$$\begin{cases}P(A+BK)+(A+BK)^{\mathrm{T}}P=-Q,\\ PB=C,\end{cases}$$

则称 $\Sigma(C,A,B)$ 为反馈正实系统. 如果 $Q>0$, 则称 $\Sigma(C,A,B)$ 为反馈严格正实系统. 在 (A,B) 为可控对且 B 为满秩矩阵时, $\Sigma(C,A,B)$ 是反馈正实系统的充分必要条件是:

1. $C^{\mathrm{T}}B>0$.
2. $\Sigma(C,A,B)$ 是弱最小相位系统.

类似地, $\Sigma(C,A,B)$ 为反馈严格正实系统的充分必要条件是:

1. $C^{\mathrm{T}}B>0$.
2. $\Sigma(C,A,B)$ 为最小相位系统.

输出反馈正实系统(output feedback positive real system) 一类线性控制系统. 如果存在矩阵 $P>0$, $Q\geqslant 0$ 和 K, 使得 $A+BKC$ 是稳定矩阵, 并且满足矩阵方程

$$\begin{cases}P(A+BK)+(A+BKC)^{\mathrm{T}}P=-Q,\\ PB=C,\end{cases}$$

则线性控制系统 $\Sigma(C,A,B)$ 称为输出反馈正实系统. 如果上面的 $Q>0$, 则称 $\Sigma(C,A,B)$ 为输出反馈严格正实系统. 设 B 是满秩矩阵, 则 $\Sigma(C,A,B)$ 为输出反馈严格正实系统的充分必要条件是:

1. $C^{\mathrm{T}}B>0$.
2. $\Sigma(C,A,B)$ 为最小相位系统.

但输出反馈正实系统至今还没有简单的判别准则.

最小相位线性系统(minimum phase linear system) 一类特殊的线性控制系统. 用 $\Sigma(C,A,B)$ 表示线性控制系统

$$\begin{cases}\dot{x}=Ax+Bu,\\ y=C^{\mathrm{T}}x,\end{cases}$$

式中 $x\in\mathrm{R}^{n}$, $u,y\in\mathrm{R}^{m}$, $A\in\mathrm{R}^{n\times n}$, $B,C\in\mathrm{R}^{n\times m}$. 设 s 是复变量. 如果当 $\mathrm{Re}(s)\geqslant 0$ 时,

$$\det\begin{pmatrix}sI-A & B\\ C^{\mathrm{T}} & 0\end{pmatrix}\neq 0,$$

那么称 $\Sigma(C,A,B)$ 为最小相位线性系统.

非线性系统

非线性控制系统理论(nonlinear control system theory) 研究非线性控制系统性质和综合方法的理论. 非线性控制系统由下列非线性向量微分方程和非线性代数方程

$$\begin{cases}\dot{x}=f(t,x,u),\\ y=h(t,x,u)\end{cases}$$

描述, 其中 t 表示时间, 在非负实数集 R_{+} 内取值 $(t\in\mathrm{R}_{+})$; x 表示状态, 在 n 维欧几里得空间 R^{n} 内变化 $(x\in\mathrm{R}^{n})$; u 表示控制输入变量, 在 r 维欧几里得空间 R^{r} 内取值 $(u\in\mathrm{R}^{r}, r\leqslant n)$, 或在 R^{r} 中某个子集合 U_{r} 内取值 $(u\in U_{r})$; y 表示量测输出变量, 在 m 维欧几里得空间 R^{m} 内变化 $(y\in\mathrm{R}^{m}, m\leqslant n)$; f,h 分别是 t,x 和 u 的非线性 n 维、m 维向量值函数. 微分方程刻画了系统的非线性动力学特征, 称为系统的状态方程. 代数方程表示系统状态、控制输入和量测输出之间的约束关系, 称为量测方程. 系统对状态和控制皆不满足叠加原理是非线性控制系统的特征, 这一特征由其状态方程和量测方程中的非线性向量值函数 $f(t,x,u)$ 和 $h(t,x,u)$ 来表示.

非线性控制系统理论是现代控制理论中发展较晚的一个分支. 非线性控制系统理论主要包括: 非线性控制系统的建模; 非线性控制系统的结构性质, 如系统的能控性、能观测性、稳定性、系统解耦和实现理论等; 非线性系统的控制性质, 如反馈镇定、调节与跟踪、动态补偿、干扰抑制和最优控制等; 非线性控制系统的设计理论, 如变结构控制方法、适应性控制和奇异摄动方法等. 20 世纪 70 年代后期以来, 采用微分几何工具研究流形上的非线性控制系统的结构、控制及相应的设计理论取得了成功. 对一类特殊非线性控制系统, 即仿射非线性系统(指控制变量线性地出现在系统的状态方程之中)已得到比较系统的结果. 特别地, 在研究通过非线性反馈和非线性坐标变换, 把定常仿射非线性控制系统局部或全局地变换成为线性定常控制系统的问题上, 即所谓反馈精确线性化和完全线性化(量测方程也变换成为线性定常代数方程)问题上, 已取得完整的理论结果. 20 世纪 80 年代中期以后, 采用微分代数方法研究非线性控制系统的结构和实现等问题, 也取得有意义的进展. 从整体上看, 在非线性控制系统理论领域内取得的结果还远不如在线性控制系统理论领域内

那样系统、完整和实用.

流形上的控制理论(control theory on manifold) 一种非线性控制系统的理论.对于非线性控制系统

$$\begin{cases}\dot{x}=f(x,u),\\ y=h(x)(\text{或 } y_i=h_i(x),i=1,2,\cdots,m);\end{cases}$$

当假定 f 是 x,u 的 n 维 C^∞ 或解析向量值函数,$h(x)=[h_1(x),h_2(x),\cdots,h_m(x)]^T$ 也是 m 维 C^∞ 或解析向量值函数时,对每个给定的常值向量 u,$f(x,\cdot)$可看成 $\mathbb{R}^n$ 上一个 C^∞ 或解析向量场,微分方程的相应轨线可看成是这个向量场的积分曲线,于是可用向量场及微分几何方法讨论非线性控制系统的性质和控制问题.几何方法中常假定所讨论非线性控制系统定义在微分流形上,即系统的状态空间是一个 n 维微分流形 M,输出空间是一个 m 维流形 N. x,y 分别是 M 和 N 上的坐标.系统的状态方程被看成是动态系统在局部坐标下的表达式,输出方程也被看成是系统输出与状态之间的关系在局部坐标下的表达式.把一个非线性控制系统定义在微分流形上,除便于使用微分几何工具外,还在于有可能讨论更一般的系统,如定义在球面、环面及一般李群上的动态系统,从而形成了一个称为流形上的控制理论的分支.

仿射非线性系统(affine nonlinear system) 一种特定的非线性系统.它对控制是线性的.当状态空间是 n 维欧几里得空间 $\mathbb{R}^n$,或者是一个一般流形时,在一个局部坐标下,系统的状态空间方程可以表示成

$$\dot{x}=f(x)+\sum_{i=1}^{r}g_i(x)u_i.$$

由于一个仿射非线性系统可以用向量场 $f(x)$,$g_i(x)(i=1,2,\cdots,r)$来完全描述,微分几何方法对这一类系统的分析显得十分有效.而且,许多实际工程系统,如机械系统等,确实属于这类系统.

双线性系统(bilinear system) 一类特定的仿射非线性系统.它的控制项对控制与状态为双线性的,其自由项对状态是线性的.因此,这类系统在适当的局部坐标下(或在 $\mathbb{R}^n$ 中)可以表达为

$$\dot{x}=Ax+\sum_{i=1}^{r}u_iB_ix.$$

双线性系统是由矩阵 $A,B_i\ (i=1,2,\cdots,r)$所决定的.因此,线性代数及矩阵群是双线性系统研究的重要工具.

无漂移系统(driftless system) 一类特殊的仿射非线性系统.即它的自由项为零.在 $\mathbb{R}^n$ 中或局部坐标下,它可以表示为

$$\dot{x}=\sum_{i=1}^{r}g_i(x)u_i.$$

这类系统有许多特点.例如,其能达集的对称性,即如果系统能由 x_1 点控制到 x_2 点,则也可以由 x_2 点控制到 x_1 点.因此,从一点出发的能达集,就是能控性李代数通过这一点的最大积分子流形.

非线性系统能控性(controllability for nonlinear system) 刻画非线性系统控制能力的一种系统结构性质.非线性系统能控性的定义较多,最常见的有以下两种:

1.强能控.系统在 x_0 点强能控指对任意一点 x,存在控制 u,使系统在 u 作用下经一段时间可以从 x_0 到达 x.

2.弱能控.系统在 x_0 弱能控指对任意一点 x,存在 $x_1,x_2,\cdots,x_k=x$,及控制 $u_i,i=1,2,\cdots,k$,使得系统在 u_i 作用下,可以从 x_{i-1}到达 x_i,或从 x_i 到达 x_{i-1}.

如果上述 x 只能取在 x_0 的某个邻域内,则可得到相应的局部能控性概念.如果系统在每一点均具有上述能控性,则称系统具有相应的能控性.

可接近性(accessibility) 刻画系统控制能力的一个结构性质.系统在 x_0 点的可达集指存在容许控制,使系统从 x_0 点出发所能达到的点的集合.如果系统在 x_0 点的可达集具有非空的内点,则称系统在 x_0 点具有可接近性;如果对任意给定的 $t>0$,系统在 t 时刻的可达集具有非空内点,则称系统在 x_0 点具有强可接近性;如果系统在每一点均具有上述可接近性,则称系统具有相应的可接近性.

分布(distribution) 向量空间中子空间概念在流形上的推广.对流形 M 上的任意一点 p,给定其切空间 T_pM 的一个子空间 $\Delta(p)$,即称在流形 M 上给定了一个分布.如果在流形上的一点 p,存在一个邻域 U 及 d 个定义在 U 上的光滑向量场 $f_1,f_2,\cdots,f_d$,使得对 U 上的每一点 p,都有

$$\Delta(p)=\mathrm{Span}\{f_1(q),f_2(q),\cdots,f_d(q)\},$$

则称该分布 Δ 在 p 点是光滑分布.如果分布 Δ 在流形上每一点都是光滑的,则称 Δ 为光滑分布.如果 Δ 在流形 M 上每点处的维数都相同,则称 Δ 为定常维数分布.非线性控制理论中经常会用到能控分布、(f,g)不变分布、受控不变分布、能控性分布等有用的分布概念.

能控分布(controllable distribution) 刻画系统能控性质的一个分布.有能控分布和强能控分布两种.对一般非线性系统

$$\dot{x}=f(x,u),$$

能控分布指的是由李代数

$$\{f(x,u)\,|\,u=\text{常数}\}_{\mathrm{LA}}$$

所决定的分布.强能控分布指的是由李代数

$$\Big\{\sum_{i=1}^{p}\lambda_iX_i+Y\,|\,p\in\mathbb{Z}^+,\sum_{i=1}^{p}\lambda_i=0\Big\}_{\mathrm{LA}}$$

所决定的分布. 这里

$$X_i \in \mathscr{F} = \{f(x,u) \mid u = \text{常数}\},$$
$$Y \in [\mathscr{F}, \mathscr{F}].$$

其中[·,·]表示李括号.

对于仿射非线性系统

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x) u_i,$$

上述概念分别对应为$\{f, g_1, g_2, \cdots, g_m\}_{LA}$和

$$\{\mathrm{ad}_f^k g_i \mid k \geqslant 0, i = 1, 2, \cdots, m\}_{LA}.$$

能控性秩条件(controllability rank condition) 用能控分布(参见"能控分布")表达的条件. 对于给定的非线性系统,当能控分布的维数在 x_0 点等于状态空间维数时,则称系统在 x_0 点满足能控性秩条件. 相应地,当强能控分布的维数在 x_0 点等于状态空间维数时,称系统满足强能控秩条件. 对于线性系统,它们都退化为能控性矩阵满秩.

非线性系统能观性(observability for nonlinear system) 刻画非线性系统输出对状态影响力的一种系统性质. 非线性系统的能观性定义较多,最常用的有以下两种:

1. 全局能观性. 全局能观性是指:如果对状态空间任意两点 x_1, x_2,都能找到一个相应的控制 u,使在该控制作用下,由初始状态 x_1 和 x_2 出发的两轨线的输出不相同.

2. 局部弱能观. 系统在 x_0 点局部弱能观系指:存在一个邻域 U,使对 U 内任一 x_0 的邻域 V, $x \in V$,都存在相应的控制 u 和时间 T,使以 x_0 与 x 出发的轨线相应的输出在时间 $0 \leqslant t \leqslant T$ 内均属于 V,且不相同. 系统是局部弱能观的,系指它在每一点均局部弱能观.

能观余分布(observability codistribution) 刻画系统能观性质的一个余分布. 对于一般控制系统

$$\begin{cases} \dot{x} = f(x,u), \\ y_j = h_j(x) \quad (j = 1, 2, \cdots, p), \end{cases}$$

能观余分布指的是余分布

$$\Omega = \mathrm{Span}\{L_{X_1}, L_{X_2}, \cdots, L_{X_k} dh_j \mid k \in \mathbf{Z}^+,$$
$$j = 1, 2, \cdots, p\},$$

这里 $X_i = f(x, u_i)$, u_i 是某些常数控制. 对于仿射非线性系统,即状态方程为

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x) u_i,$$

则相应的 X_i 为 $f(x)$或 $g_j(x)$ $(j=1,2,\cdots,m)$.

能观性秩条件(observability rank condition) 用能观余分布(参见"能观余分布")的维数描述的一个能观性条件. 如果能观余分布的维数与状态空间的维数相等,则称系统满足能观性秩条件. 对于线性系统,能观性秩条件等价于能观性矩阵满秩.

标准分解(canonical decomposition) 系统在特殊坐标下的一种特殊形式. 它指的是非线性系统经过适当的坐标变换后可以变为依某种性质分解开的标准形式,主要有以下几种:

1. 能控分解. 系统经适当坐标变换后使系统空间分解成两部分. 一部分满足能控性秩条件,而另一部分则完全不能控.

2. 能观分解. 系统经坐标变换后使系统空间分解成两部分. 一部分满足能观性条件,而另一部分则完全不能观.

3. 卡尔曼分解. 系统空间经坐标变换后可分解成四部分,分别对应能控能观、能控不能观、不能控能观、不能控不能观的状态.

(f,g)不变分布((f,g)-invariant distribution) 一种特殊的分布. 对于仿射非线性系统

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x) u_i = f(x) + g(x) u,$$

如果存在一个分布 Δ,以及反馈

$$u = \alpha(x) + \beta(x) v,$$

这里$\beta(x)$为非奇异函数矩阵,使得对

$$\tilde{f}(x) = f(x) + g(x)\alpha(x),$$
$$\tilde{g}(x) = g(x)\beta(x)$$

成立

$$[\tilde{f}, \Delta] \subset \Delta,$$
$$[\tilde{g}_i, \Delta] \subset \Delta \quad (i = 1, 2, \cdots, m),$$

则称 Δ 为(f,g)不变分布. 如果

$$[f, \Delta] \subset \Delta + G,$$
$$[g_i, \Delta] \subset \Delta + G \quad (i = 1, 2, \cdots, m),$$

这里 $G = \mathrm{Span}\{g_1, \cdots, g_m\}$,则称 Δ 为弱(f,g)不变分布. 它们的关系可参见"奎克引理".

受控不变分布(controlled invariant distribution) 控制系统的一种特殊的分布. 对于一般非线性系统 $\dot{x} = f(x,u)$,如果存在一个反馈控制 $u = u(x,v)$,使得对每一个给定的 x, $v \mapsto u$ 是一个同胚映射,则称 $u = u(x,v)$是一个正则反馈. 如果存在一个正则反馈 $u(x,v)$,使得对每一个定常的 v 有 $[f(x, u(x,v)), \Delta] \subset \Delta$,则称分布 Δ 为能控不变分布. 在仿射非线性系统的情况下,能控不变分布退化为(f,g)不变分布.

能控性分布(controllability distribution) 仿射非线性系统的一种特殊的(f,g)不变分布. 对仿射非线性系统

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x) u_i = f + gu,$$

如果存在反馈 $u = \alpha(x) + \beta(x) v$ ($\beta(x)$可逆),以及$\{1, 2, \cdots, m\}$的一个子集 A,使得

$$R = \langle \tilde{f}, \tilde{g}_1, \tilde{g}_2, \cdots, \tilde{g}_m \mid g^A \rangle,$$

即 R 为包含 g^A,且对 $\tilde{f}, \tilde{g}_1, \tilde{g}_2, \cdots, \tilde{g}_m$ 不变的最小分

布，则称 R 为一个能控性子分布. 这里

$$\tilde{f}=f+g\alpha,\ \tilde{g}=g\beta.$$

相容(f,g)不变分布族(compatible (f,g)-invariant distribution) 一族特殊的(f,g)不变分布. 一族线性无关的非奇异对合分布 $\Delta_1,\Delta_2,\cdots,\Delta_k$，如果对仿射非线性系统

$$\dot{x}=f(x)+\sum_{i=1}^{m}g_i(x)u_i=f+gu$$

存在同一个反馈 $u=\alpha(x)+\beta(x)v$，使得对反馈系统 $\dot{x}=f+g\alpha+g\beta v=\tilde{f}+\tilde{g}v$ 每个 Δ_i 均不变，即

$[\tilde{f},\Delta_i]\subset\Delta_i$,

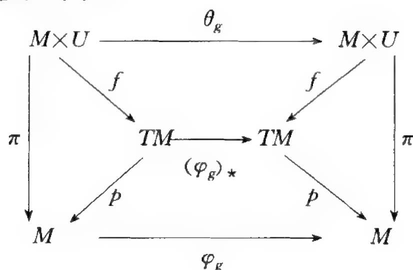
$[\tilde{g}_j,\Delta_i]\subset\Delta_i\ (j=1,2,\cdots,m,i=1,2,\cdots,k)$,

则 $\Delta_1,\Delta_2,\cdots,\Delta_k$ 称为相容(f,g)不变分布族.

对称系统(symmetric system) 一种特殊类型的非线性控制系统. 一般非线性系统

$$\dot{x}=f(x,u)\ (x\in M,u\in U)$$

可以用纤维丛框架描述. 对称性在纤维丛框架下可定义如下：如果存在一个群 G，它在状态空间 M 及状态-输入空间 $M\times U$ 各有一个群作用 φ_g 与 θ_g，使得图中所示交换成立，则称系统在群 G 作用下对称. 如果 $\theta_g=\varphi_g\times id_v$，则称系统在群 G 作用下状态对称. 可参见下图.



无穷小对称(infinitesimal symmetry) 描述系统对称结构的一个概念. 设系统

$$\dot{x}=f(x,u),\quad x\in M\quad u\in U,$$

有群作用 G：

$$\theta_g:M\times U\to M\times U,$$

$$\varphi_g:M\to M,\quad g\in G,$$

记 G 的李代数为 $\mathscr{J}$，如果对任一 $x\in\mathscr{J}$，$\|x\|<1$，存在 $\varepsilon>0$，使得系统对 $\theta_{\exp(t)\cdot x}$ 及 $\varphi_{\exp(t)\cdot x}(0<t<\varepsilon)$ 对称(参见"对称系统")，则称系统在 G 作用下无穷小对称.

线性化(linearization) 将非线性系统转化为线性系统的一种方法. 主要有局部近似线性化和精确线性化两种. 对于非线性控制系统，若(x_0,u_0)是 $f(x,u)$的驻点，令

$$A=\left.\frac{\partial f}{\partial x}\right|_{(x_0,u_0)},\ B=\left.\frac{\partial f}{\partial u}\right|_{(x_0,u_0)},$$

那么线性控制系统 $\dot{z}=Az+Bu$ 称为其在(x_0,u_0)点的(局部)近似线性化系统. 这种线性化(方法)称为经典线性化(方法). 一个非线性控制系统，若存在局部坐标系(φ,U)(或非奇异状态变换 $z=\varphi(x)$)，使得在新的状态坐标下，它具有线性控制系统的形式，则称其为可(局部)精确线性化的. 若可取 $U=\mathbb{R}^n$，n 为非线性控制系统状态的维数，则称其为是可(全局)精确线性化的.

反馈线性化(feedback linearization) 基于精确线性化思想所发展的一种反馈设计方法. 一个非线性控制系统，若存在状态反馈，使得状态反馈作用后的闭环系统是可(局部或全局)精确线性化的，则称其为可(局部或全局)反馈线性化的. 对于能反馈线性化的非线性系统，可在线性化后再按线性系统的设计方法设计控制规律来达到预期的目的. 仿射非线性控制系统(参见"仿射非线性系统")可(局部)反馈线性化，且成为一个能控的线性控制系统的充分必要条件是：

1. 对于 $i=0,1,\cdots,n-1$,

$$\Delta_i=\mathrm{Span}\{g_1,\cdots,g_m,\mathrm{ad}_fg_1,\cdots,\mathrm{ad}_f^ig_1,\cdots,\mathrm{ad}_f^ig_m\}$$

是对合分布.

2. Δ_i 在局部具有常维数.

3. $\dim\Delta_{n-1}=n$.

输入-输出线性化(input-output linearization) 非线性系统的一种控制设计方法. 一个仿射非线性控制系统，如果在其沃尔特拉展式中，二阶及二阶以上的高阶核为零，则称其为具有线性输入-输出关系. 如果存在一个状态反馈，使得状态反馈作用后的闭环系统具有线性输入-输出关系，则称其为可(反馈)输入-输出线性化. 记 $T_k(x)$为 $p\times m$ 矩阵，其第(i,j)个元素为 $L_{g_j}L_f^kh_i(x)$. 定义一个托埃普里茨(Toeplitz)矩阵序列 $M_k(x)\ (k\geqslant0)$,

$$M_k(x)=\begin{bmatrix}T_0(x)&T_1(x)&\cdots&T_k(x)\\0&T_0(x)&\cdots&T_{k-1}(x)\\\vdots&\vdots&&\vdots\\0&0&\cdots&T_0(x)\end{bmatrix}.$$

设 x_0 是 $M(x)$的正则点，其秩记为$r_{\mathrm{R}}(M_k)$，而记 $r_{1k}(M_k)$为 M 作为实数域 R 上矩阵的秩. 那么在 x_0 点局部，仿射非线性控制系统可输入-输出线性化的充分必要条件是：对所有 $k\geqslant0$，$r_{\mathrm{R}}(M_k)=r_{1k}(M_k)$.

输入-输出解耦(input-output decoupling) 一个典型的控制设计问题. 考虑非线性控制系统

$$\dot{x}=f(x)+\sum_{i=1}^{m}g_i(x)u_i=f(x)+g(x)u,$$

$$y^i=h^i(x)\ (i=1,2,\cdots,p),$$

其中输出 $y\in\mathbb{R}^p$ 被分成了 k 组,

$$y^1,y^2,\cdots,y^k,y^i\in\mathbb{R}^{p_i}\ (i=1,2,\cdots,k),$$

$$\sum_{i=1}^{m}p_i=p.$$

如果存在输入 u 的一个分组

$$u^{\mathrm{T}} = (u^1, u^2, \cdots, u^k, u^{k+1}),\ u^i \in \mathbb{R}^{m_i},$$
$$\sum_{i=1}^{k+1} m_i = m,$$

使得 u^i 只影响 y^i，而不影响 $y^j(i,j=1,2,\cdots,k,j\neq i)$，$u^{k+1}$对所有输出都没有影响，则称其为输入-输出解耦的. 如果存在状态反馈，使得状态反馈作用后的闭环系统为输入-输出解耦的，则称系统为可反馈解耦的. 而寻找这个状态反馈使系统成为解耦形式的问题则称为(输入-输出)(块)解耦问题. 特别地，如果 $k=m, p_i=1(i=1,2,\cdots,k)$，则称为单对单解耦问题，或称为(非线性)摩根问题.

特征指数(characteristic number)　描述输出和输入相关关系的一种度量. 设 $y_i=h_i(x)$为非线性控制系统的第 i 个输出分量，若整数 ρ_i 满足性质：

1. $L_{g_j}L_f^k h_i(x)=0, k<\rho_i, 1\leqslant j\leqslant m$；
2. 存在某个 j，使得 $L_{g_j}L_f^{\rho_i}h_i(x)\neq 0$；

则称 ρ_i 为系统第 i 个(输出 y_i 的)特征指数(参见"相对阶").

解耦矩阵(decoupling matrix)　描述输入-输出耦合关系的一个矩阵. 设 ρ_i 是非线性控制系统第 $i(i=1,2,\cdots,p)$个特征指数，又设每个 ρ_i 都是有限的. 记

$$A(x)=\begin{bmatrix} L_f^{\rho_1+1}h_1(x) \\ \vdots \\ L_f^{\rho_p+1}h_1(x) \end{bmatrix},$$

$$B(x)=\begin{bmatrix} L_{g_1}L_f^{\rho_1}h_1(x) & \cdots & L_{g_m}L_f^{\rho_1}h_1(x) \\ \vdots & & \vdots \\ L_{g_1}L_f^{\rho_p}h_p(x) & \cdots & L_{g_m}L_f^{\rho_p}h_p(x) \end{bmatrix},$$

则矩阵 $B(x)$称为系统的解耦矩阵，亦称为法尔布-沃洛维奇矩阵. 此时，系统的输出和输入满足下面的关系：

$$\begin{bmatrix} y_1^{(\rho_1+1)} \\ \vdots \\ y_p^{(\rho_p+1)} \end{bmatrix} = A(x) + B(x)u.$$

解耦条件(decoupling condition)　控制系统能实现反馈解耦控制的条件. 摩根问题(通过正则状态反馈)局部可解的充分必要条件是解耦矩阵 $B(x)$是局部行满秩的. 记 Δ_i^* 为包含在 $\ker dh^i(x)$中($h^i(x)$为第 i 组输出)最大的(f,g)不变分布，R_i^* 为包含在 $\ker dh^i(x)$中最大能控性分布. 假设 Δ_i^* 对每个 i 都存在. 如果非线性系统的(块)解耦问题可解，那么

$$\Delta_i^* \cap G + \bigcap_{j\neq i}(\Delta_j^* \cap G) = G \quad (i=1,2,\cdots,k), \quad (1)$$

其中 $G=\mathrm{Span}\{g_1,g_2,\cdots,g_m\}$. 反过来，如果 $\Delta_i^*, G, \Delta_i^*\cap G, \bigcap_{j\neq i}(\Delta_j^*\cap G)(i,j=1,2,\cdots,k)$具有常维数，且(1)式成立，则(块)解耦问题局部可解. 类似地，假设 R_i^* 对每个 i 都存在，如果非线性系统的(块)解耦问题可解，那么

$$R_i^* \cap G + \bigcap_{j\neq i}(R_j^* \cap G) = G \quad (i=1,2,\cdots,k). \quad (2)$$

反过来，如果 $R_i^*, G, R_i^*\cap G, \bigcap_{j\neq i}(R_j^*\cap G)(i,j=1,\cdots,k)$具有常维数，且(2)式成立，则(块)解耦问题局部可解.

奎克引理(Quaker lemma)　判断一个给定分布是否是(f,g)不变分布的一种方法.

奎克引理：设 Δ 是一个局部非奇异对合分布，并且 $G, G\cap\Delta$ 也是局部非奇异的，则 Δ 是局部(f,g)不变分布的充分必要条件是：

$$[f,\Delta]\subset \Delta + G, [g_i,\Delta]\subset \Delta + G$$
$$(i=1,2,\cdots,m).$$

奎克引理中关于 $G, \Delta\cap G$ 的正则性条件也可减弱为G/Δ 是局部非奇异的，这里G/Δ 为 G 相对于 Δ 的投影.

沃尔特拉展式(Volterra expansion)　用特殊形式的多重积分之和表达的一种非线性输入输出关系. 令 $S_k=\{(t,\tau_k,\tau_{k-1},\cdots,\tau_1)\in\mathbb{R}^{k+1}; T\geqslant t\geqslant\tau_k\geqslant\cdots\geqslant\tau_1\geqslant 0\}, k=0,1,\cdots$. 设 $W_{i_k i_{k-1}\cdots i_1}$为定义在 S_k 上的实值连续函数，如下形式的多重卷积序列和称为沃尔特拉级数：

$$y(t)=W_0(t)+\sum_{k=1}^{\infty}\sum_{i_1,\cdots,i_k=1}^{m}\int_0^t \mathrm{d}\tau_k\int_0^{\tau_k}\mathrm{d}\tau_{k-1}\cdots$$
$$\cdot\int_0^{\tau_3}\mathrm{d}\tau_2\int_0^{\tau_2}[W_{i_k\cdots i_1}(t,\tau_k,\tau_{k-1},\cdots,\tau_1)$$
$$\cdot u_{i_k}(\tau_k)\cdots u_{i_1}(\tau_1)]\mathrm{d}\tau_1.$$

如果一个仿射非线性控制系统的每个输出分量都具有如上的沃尔特拉级数展开，则称其具有沃尔特拉展式描述. 对于仿射非线性控制系统，如果 $f, g_1, g_2, \cdots, g_m$ 是解析向量场，每个输出分量映射 $h_i(x)$是解析函数，设

$$\dot{x}=f(x),\ x(0)=x_0$$

在$[0,T]$上存在一个解，那么系统对于 $t\in[0,T]$具有沃尔特拉展式描述，且是惟一的.

在沃尔特拉展式描述中，$W_{i_k i_{k-1}\cdots i_1}$称为第 $i_k, i_{k-1},\cdots,i_1$ 个 k 重核.

弗里斯展式(Fliess expansion)　描述非线性输入输出关系的一种无穷和展开式. 先递推定义一族积分：

$$\int_0^t \mathrm{d}\xi_i=\int_0^t u_i(\tau)\mathrm{d}\tau \quad (i=0,1,\cdots,m),$$
$$\int_0^t \mathrm{d}\xi_{i_k+1}\mathrm{d}\xi_{i_k}\cdots\mathrm{d}\xi_{i_1}=\int_0^t u_{i_k+1}(\tau)\int_0^\tau u_{i_k}\cdots\mathrm{d}\xi_{i_1}(k\geqslant 1),$$

那么，如下形式的多重积分序列和称为弗里斯级数：

$$y(t)=h(x_0)+\sum_{k=0}^{\infty}\sum_{i_0,\cdots,i_k=0}^{m}[L_{g_{i_0}}\cdots L_{g_{i_k}}h(x_0)$$

$\cdot \int_0^t d\xi_{i_k} \cdots d\xi_{i_0}]$.

如果一个仿射非线性控制系统的每个输出分量都具有如上的弗里斯级数展开，则称其具有弗里斯展式描述. 对于仿射非线性控制系统，如果 $f, g_1, g_2, \cdots, g_m$ 是解析向量场，每个输出分量映射 $h_i(x)$ 是解析函数，设

$$\dot{x} = f(x),\ x(0) = x.$$

在$[0,T]$上存在一个解，那么系统对于 $t \in [0,T]$ 具有弗里斯展式描述，且是惟一的，其中令 $g_0 = f$，$u_0(t) = 1$. 一个仿射非线性控制系统的沃尔特拉展式描述和弗里斯展式描述间有下列关系：

$$W_0(t) = \sum_{n=0}^{\infty} L_f^n h_j(x_0) \frac{t^n}{n!},$$

$$W_{i_1}(t,\tau_1) = \sum_{n_1=0}^{\infty}\sum_{n_0=0}^{\infty} L_f^{n_0} L_{g_i} L_f^{n_1} h_j(x_0) \frac{(t-\tau_1)^{n_1}}{n_1!}\frac{\tau_1^{n_0}}{n_0!},$$

$$W_{i_2 i_1}(t,\tau_2,\tau_1) = \sum_{n_2=0}^{\infty}\sum_{n_1=0}^{\infty}\sum_{n_0=0}^{\infty}\Big[L_f^{n_0} L_{g_{i_1}} L_f^{n_1} L_{g_{i_2}} L_f^{n_2} \cdot h_j(x_0) \frac{(t-\tau_2)^{n_2}(\tau_2-\tau_1)^{n_1}\tau_1^{n_0}}{n_2!\ n_1!\ n_0!}\Big]$$

等. 弗里斯展式是弗里斯(Fliess, M.)提出的.

实现(realization)　描述已知输入输出关系的状态空间方程. 对于给定的一个输入输出之间的数据序列(一种输入输出关系的表达)，例如给定一个收敛的沃尔特拉级数，或一个收敛的弗里斯级数，如果存在以状态方程和输出方程描述的非线性控制系统，其相应的输入输出数据(描述)，例如其沃尔特拉展式描述或其弗里斯展式描述存在，且等于给定的输入输出数据，则称这个状态空间描述为给定输入输出数据的一个实现.

寻找这个实现所构成的问题称为实现问题. 一般情况下的实现问题尚未完全解决.

仿射实现(affine realization)　描述已知输入输出关系的状态方程为仿射型的一种特殊实现. 如果在实现问题中，寻找的是一个由仿射非线性控制系统方程所表达的状态空间描述，则称这个仿射非线性控制系统为给定输入输出数据序列的一个仿射实现. 特别地，如果这个仿射非线性控制系统是如下形式的双线性控制系统

$$\dot{x} = Ax + \sum_{i=1}^{m}(B_i x + D_i)u_i,\ y = Cx,$$

则称这时的实现为双线性实现. 设给定的输入输出数据是由一个收敛的弗里斯级数给出的，或等价地，选择 $z_0, z_1, \cdots, z_m$ 为未定元，设给定如下形式的幂级数：

$$C(z) = C + \sum_{k=0}^{\infty}\sum_{i_0,\cdots,i_k=0}^{m} C(i_0,i_1,\cdots,i_k) z_{i_k} z_{i_{k-1}} \cdots z_{i_0} \in R^p\langle z\rangle,$$

其中单项式 $z_{i_k}, z_{i_{k-1}}, \cdots, z_{i_0}$ 的因子 $z_0, z_1, \cdots, z_k$ 不可交换，$C(i_0,i_1,\cdots,i_k) \in R^p\langle z\rangle$，$R^p\langle z\rangle$ 为全体形式幂级数集合所构成的线性空间. 由有限形式幂级数(即只有有限项系数不为零)全体所构成的线性空间记为 $R_0^r\langle z\rangle$. $R_0^1\langle z\rangle$ 在李括号 $[p,q] = pq - qp$ 下成为李代数. 记 $L(z)$ 为包含 $z_0, z_1, \cdots, z_m$ 的最小子代数. 定义线性映射 $F_C: R_0^1\langle z\rangle \to R^r\langle z\rangle$，

$$\begin{cases} [F_C(1)](i_0,i_1,\cdots,i_r) = C(i_0,i_1,\cdots,i_r), \\ [F_C(z_{j_k}, z_{i_{k-1}}, \cdots, z_{j_0})](i_0,i_1,\cdots,i_r) \\ = C(j_0,j_1,\cdots,j_k,i_0,i_1,\cdots,i_r). \end{cases}$$

$F_C(R_0^1\langle z\rangle)$ 及 $F_C(L(z))$ 的维数分别称为 F_C 的汉克尔秩和李秩，且分别记为 $\rho_H(C)$ 和 $\rho_L(C)$.

$C(z)$ 具有双线性实现，当且仅当 $\rho_H < \infty$.

$C(z)$ 具有仿射实现，当且仅当 $\rho_L < \infty$.

最小实现(minimal realization)　其状态方程维数最小的一种实现. 一个实现中状态(方程)的维数称为该实现的维数. 如果某一个实现的维数小于或等于其他实现的维数，则称该实现为最小实现. 形式幂级数 $C(z)$ 的一个实现是最小实现，当且仅当该实现的维数等于 F_C 的李秩 $P_L(C)$；形式幂级数 $C(z)$ 的一个实现是最小实现，当且仅当该实现满足能控性秩条件及能观性秩条件. 形式幂级数 $C(z)$ 的最小实现在局部微分同胚意义下是惟一的.

微分代数方法(differential algebraic method)　利用微分代数为数学工具来研究控制系统的一种方法. 它的主要思想如下：设 K 是给定的一个微分域，有限生成的微分域扩张 $K\langle y,u\rangle/K$，如果还满足

$$\text{diff. tr. } d^0 K\langle y,u\rangle/K\langle u\rangle = 0,$$

即微分域扩张 $K\langle y,u\rangle/K\langle u\rangle$ 是微分代数的，则称其为一个输入为 u、输出为 y 的非线性控制系统. 设 $(x_1, x_2, \cdots, x_n) = x$ 为域扩张 $K\langle y,u\rangle/K\langle u\rangle$ 的超越基，那么，非线性控制系统 $K\langle y,u\rangle/K$ 具有如下广义状态方程描述(实现)：

$$f(\dot{x}, x, u, \dot{u}, \cdots, u^{(\alpha)}) = 0,$$
$$h(y, x, u, \dot{u}, \cdots, u^{(\alpha)}) = 0.$$

微分代数方程所定义的非线性系统较通常的状态空间方程描述的系统要广泛，研究内容也要丰富. 微分代数方法主要是由法国控制数学家弗里斯(Fliess, M.)发展起来的. 这个方法尚在发展之中.

微分输出秩(differential output rank)　非线性控制系统的微分代数方法中为描述输入输出关系的一个量. 对于给定的非线性控制系统 $K\langle y,u\rangle/K$，整数 $\rho = \text{diff. tr. } d^0 K\langle y\rangle/K$ 称为其微分输出秩. 对于通常的线性时不变控制系统，其微分输出秩等于其传递函数矩阵的秩.

相对阶(relative degree)　输入与输出之间相关性的一种度量. 考虑输入和输出维数相等的仿射

系统

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x)u_i,$$
$$y_i = h_i(x) \ (i = 1,2,\cdots,m),$$

其中 x 是 n 维状态变量. 如果在 x_0 的某个邻域 U 内,对一切 $i=1,2,\cdots,m$ 和一切 $k<r_i-1$,成立

$$L_{g_i}L_f^k h_i(x)=0, \quad \forall x\in U,$$
$$L_{g_{i_0}}L_f^{r_i-1}h_i(x)=0, \ i_0\in\{1,2,\cdots,m\},$$

并且矩阵

$$\begin{bmatrix} L_{g_1}L_f^{r_1-1}h_1(x) & L_{g_2}L_f^{r_1-1}h_1(x) & \cdots & L_{g_m}L_f^{r_1-1}h_1(x) \\ L_{g_1}L_f^{r_2-1}h_1(x) & L_{g_2}L_f^{r_2-1}h_1(x) & \cdots & L_{g_m}L_f^{r_2-1}h_2(x) \\ \vdots & \vdots & & \vdots \\ L_{g_1}L_f^{r_m-1}h_1(x) & L_{g_2}L_f^{r_m-1}h_1(x) & \cdots & L_{g_m}L_f^{r_m-1}h_m(x) \end{bmatrix}$$

在 x_0 处是非奇异的,那么称系统在 x_0 处有相对阶 $(r_1,r_2,\cdots,r_m)$. 对于一般多变量仿射系统,未必在每点都存在相对阶.

零动态(zero dynamics) 线性系统零点概念在非线性系统中的推广. 对于系统

$$\dot{x}_1 = f_1(x_1,u),$$
$$\dot{x}_2 = f_2(x_1,x_2,u),$$
$$y = h(x_1),$$

其中

$$x=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

是状态,u 和 y 分别是输入和输出. 如果 $y(t)\equiv 0$ 和 x_1 的初态 $x_1^0=0$ 总有惟一的输入反馈 $u=u^*(x)$,使得 $x_1(t)\equiv 0$,那么 $\dot{x}_2\equiv f_2(0,x_2,u^*)$ 称为该系统的零动态(系统). 零动态是在研究高增益反馈时提出的,本斯(Byrnes,C.)和依斯朵里(Isidori,A.)曾经提出一个零动态算法来得到仿射非线性系统的零动态. 零动态是线性系统传递零点在非线性系统中的推广,它刻画了输出等于零时,系统内部的运动情况.

干扰解耦(disturbance decoupling) 设计状态反馈使干扰不影响闭环系统输出的一个典型的控制设计问题. 对非线性系统

$$\dot{x} = f(x,u,w),$$
$$y = h(x,u,w),$$

其中 u,x,y 和 w 分别是输入、状态、输出和干扰,$y(t,x_0,u,w)$ 表示初态为 x_0、输入和干扰分别为 u 和 w 的输出轨线. 如果对任意两个 $w_1\neq w_2$ 都有

$$y(t,x_0,u,w_1) = y(t,x_0,u,w_2),$$

则称为是干扰解耦的. 如果只成立

$$y(t,x_0,u,w_1) - y(t,x_0,u,w_2) \to 0(t\to\infty),$$

那么称为是几乎干扰解耦的. 仿射系统

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x)u_i + p(x)w$$

的干扰解耦有完善的结论:上述系统能用状态反馈实现干扰解耦的充分必要条件是:存在一个 (f,g_i) 不变分布 Δ,使得 $p(x)\in\Delta(x)\subset\ker dh(x)$. 上式只在局部成立就是局部干扰解耦,在全局成立便是全局干扰解耦.

最小阶补偿器的解耦控制(minimal order compensator for decouped control) 一种阶数最小的动态解耦控制. 对于一般非线性系统,采用

$$\dot{\xi} = \alpha(\xi,u,y),$$
$$u = \beta(\xi,y,v)$$

实现反馈,称为动态补偿器. 其中 v 是参考输入. ξ 的维数称为补偿器的阶,阶数最低的补偿器称为最小阶补偿器. 根据不同的目的可以有不同的最小阶补偿控制,例如最小阶补偿的输入-输出解耦(摩根问题)控制、最小阶补偿的干扰解耦控制. 由于动态补偿是比状态反馈更广泛的控制形式,所以作用也更大. 通过对最小阶补偿器的解耦控制问题的研究也能够揭示许多深刻的系统结构性质.

李群上的右不变系统(right-invariant system on Lie group) 一类特殊的非线性控制系统. G 是一个李群,$g\in G$,用 R_g 表示 G 上的一个右变换. G 上的李代数记为 $\mathscr{G}$,R_{g*} 则为 $\mathscr{G}$ 上的右推移算子,即 f 为 G 上的向量场 $R_{g*}f(a)=f(ga)$. 如果对一切 $g\in G$ 都成立 $R_{g*}f=f$,则 f 称为是右不变的. 如果 f 和 $g_i(i=1,2,\cdots,m)$ 都是右不变向量场,那么就称李群 G 上的仿射系统

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x)u_i,$$

其中 $x\in G,u_i\in\Omega$(Ω 为某个容许输入集合)是一个右不变系统. 李群上的右不变系统有较好的可达性质. 例如,用 $WA(g)$ 表示 $g\in G$ 的弱可达集,那么 $WA(g)=WA(e)g$,其中 e 是李群上的单位元. 用 $A(e)$ 表示 e 的可接近集合,$\mathring{A}(e)$ 是 $A(e)$ 的内点,则

$$\dim\mathring{A}(e) = \dim W\mathring{A}(e),$$

而且 $WA(e)$ 是 G 的一个李子群.

非线性系统的可逆性(invertibility of nonlinear system) 描述系统输入和输出映射的单射性和满射性概念. 是由映射的可逆性演化而来的. Σ_1 和 Σ_2 是两个非线性系统,Σ_1 的输出维数等于 Σ_2 的输入维数,x_1^0 和 x_2^0 分别是 Σ_1 和 Σ_2 的状态变量的初值,u_i 和 y_i 是 $\Sigma_i(i=1,2)$ 的输入和输出. 如果对任意的 x_1^0 都存在 x_2^0,使得在这时 $\Sigma_1\circ\Sigma_2 y_1=y_1$,则称 Σ_1 是右可逆的,Σ_2 是 Σ_1 的一个右逆系统.

类似地,如果对一切 x_1^0,使得 $\Sigma_1\circ\Sigma_2 u_2=u_2$,则称 Σ_2 是左可逆的,Σ_1 为 Σ_2 的一个左逆系统. 可逆性在控制系统设计中是一个重要的性质.

逆系统(inverse system) 一种与给定系统具有某种相反的输入输出传递关系的系统.从函数映射的观点看,系统是一个由输入函数集合 $u(\cdot)\in U$ 到输出函数集合 $y(\cdot)\in Y$ 在给定初始条件 $x_0\in X$ 下的映射 $\theta:u\to y$,或简化为 $y=\theta u$.而逆系统则是由该映射 θ 所定义的逆映射 $\hat{\theta}:y\to u$,或简记为 $u=\hat{\theta}y$.在逆系统的有关概念中,左逆系统和右逆系统是两种最基本的分类方式.左可逆性表示可由系统的输出信号 $y(\cdot)$ 恢复出原来的输入信号 $u(\cdot)$,即左逆系统 $\hat{\theta}$ 满足 $\hat{\theta}\theta u(\cdot)=u(\cdot)$;右可逆性表示对于某些给定的信号 $y(\cdot)$,可以找到相应的信号 $u(\cdot)$ 作为系统的输入,使系统在这种输入信号的激励下,以给定的信号 $y(\cdot)$ 为输出,即右逆系统满足 $\theta\hat{\theta}y(\cdot)=y(\cdot)$.对于由方程

$$\dot{x}=f(x,u),\ y=h(x,u)$$

所表示的多输入-多输出一般非线性系统,其中 u 为输入,y 为输出.如果该系统是可逆的,其逆系统方程可通过一种基于函数求导和隐函数定理的一般结构算法导出.逆系统方程的一般形式为

$$\dot{x}=f(x,h_a^{-1}(x,y,y',\cdots,y^{(\alpha)})),$$
$$u=h_a^{-1}(x,y,y',\cdots,y^{(\alpha)}),$$

其中 y 为输入,u 为输出,正整数 α 被规定为所给系统的相对阶数,它是一个重要的不变量.例如,上述一般非线性系统的左和右可逆性的充分必要条件可统一表述为 $\alpha\leqslant n=\dim x$.逆系统理论在控制理论的许多方面都有广泛的应用,如解耦、模型匹配、最小设计、多变量线性系统零点的定义与求解、多变量非线性系统设计的逆系统方法、编码理论、滤波、对策论等方面.

逆系统方法(inverse system method) 基于动态系统可逆性质设计非线性系统控制器的一般理论方法.逆系统方法的设计思想是:由被控系统方程出发,首先设计出被控对象的 α 阶积分逆系统方程,将它作为控制律对系统实行非线性反馈控制,从而设计出具有线性传递关系的伪线性系统,然后再用线性系统理论来完成控制系统设计.如下图所示.

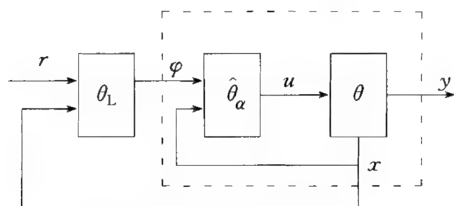


图1

设被控系统 θ 为一般的多输入-多输出非线性系统,其方程为

$$\dot{x}=f(x,u),\quad y=h(x,u).$$

其 α 阶积分逆系统 $\hat{\theta}_\alpha$ 的方程为

$$\hat{\theta}_\alpha:\begin{cases}\dot{\xi}=L_1\xi+L_2\varphi,\\ u=K(x,\xi,\varphi).\end{cases}$$

系统经过由 α 阶积分逆系统 $\hat{\theta}_\alpha$ 的方程所定义的控制律反馈控制后,形成伪逆系统,它满足方程

$$y_i^{(\alpha)}=\varphi_i\ (i=1,2,\cdots,r),\ r=\dim y.$$

然后,对伪逆系统作用线性控制器 θ_L 即可构成完整的控制系统,它具有从参考输入 r 到输出 y 的某种预期的控制特性,如实现解耦、镇定、特征结构配置、二次型最优、鲁棒伺服跟踪等控制系统设计目标.

模型匹配(model matching) 一个典型的控制器设计问题.给定非线性系统 Σ 和模型系统 Σ_m,x^0 和 x_m^0 分别是 Σ 和 Σ_m 的状态初值.$y(t,x^0,u)$ 和 $y_m(t,x_m^0,u)$ 分别是 Σ 和 Σ_m 在输入 u 和初态 x^0 与 x_m^0 时的输出轨线.模型匹配问题就是对 Σ 寻找反馈律使得对一切容许输入 u 和一切 x_m^0 与 x^0,都成立

$$y_m(t,x_m^0,u)=y(t,x^0,u).$$

对于仿射非线性系统来说,实现模型匹配的关键在于 Σ 和 Σ_m 有相同的无穷远结构.

离散时间非线性系统(discrete-time nonlinear system) 在离散的时间点上描述的非线性控制系统.形式为

$$x(k+1)=f(x(k),u(k)),$$
$$y(k)=h(x(k),u(k))$$

的控制系统称为离散时间系统.当 f 或者 h 是非线性算子时,则称上述系统为离散时间非线性系统.与线性系统理论比较,离散时间非线性系统和连续时间非线性系统之间的差别要大得多.

非线性系统观测器(observers for nonlinear system) 用非线性测量以实现非线性系统的状态估计的算法或装置.对于非线性系统

$$\begin{cases}\dot{x}=f(x,u),\\ y=h(x,u),\end{cases}\tag{1}$$

存在另一个系统

$$\dot{\tilde{x}}=\mathcal{F}(\tilde{x},u,y),\tag{2}$$

使得 $x-\tilde{x}\to0(t\to\infty)$ 对一切初始状态 x_0 和 $\tilde{x}_0$ 都成立,那么称(2)是(1)的一个(渐近)状态观测器.如果当 $x_0=\tilde{x}_0$ 时还成立 $x(t)\equiv\tilde{x}(t)$,则称(2)是(1)的状态观测器.非线性系统的能观性不直接蕴含观测器存在,从而观测器成了一个独立的研究课题.已经提出的方法有李代数方法、李亚普诺夫方法、扩展线性化方法、浸入方法和变结构方法等.每种设计方法都有各自的特点和局限,尚缺乏一般设计方法.

李亚普诺夫稳定性(Liapunov stability) 一种用于研究自由系统初态受扰后轨线渐近性质的概念.不受控系统

$$\dot{x}=f(x,t)\tag{1}$$

称为自由非线性系统,其中 $x\in\mathrm{R}^n$,t 表示时间,$f(x,$

t)为定义在$\Omega\times[0,+\infty)$上的向量函数,Ω为R^n中某个区域,Ω常常是R^n本身,R^n为n维欧几里得空间(简称欧氏空间).假定(1)在t_0时刻以$x_0(\in\Omega)$为初值的解$x(t)\triangleq x(t;t_0,x_0)$在$[t_0,+\infty)$上存在、惟一.用$\|\cdot\|$表示·的欧几里得范数.称$x_e$(常向量)为系统(1)的平衡状态,如果$x_e$是向量微分方程

$$f(x_e,t)=0 \tag{2}$$

的解.李亚普诺夫稳定、渐近稳定、不稳定、一致稳定和一致渐近稳定的定义分别为:

1. 如果对任给的$\varepsilon>0$,可以找到一个与ε和t_0有关的$\delta=\delta(\varepsilon,t_0)$,使得当

$$\|x_0-x_e\|\leqslant\delta \tag{3}$$

时,对一切$t\geqslant t_0$,恒成立

$$\|x(t;t_0,x_0)-x_e\|<\varepsilon, \tag{4}$$

则称平衡状态x_e是(局部)稳定的.

2. 如果平衡状态x_e是稳定的,而且有

$$\lim_{t\to\infty}\|x(t;t_0,x_0)-x_e\|=0, \tag{5}$$

则称平衡状态x_e是(局部)渐近稳定的.

3. 如果存在正数ε,使满足1中要求的δ不存在,则称平衡状态x_e是不稳定的.

4. 当1和2中的$\delta(\varepsilon,t_0)$仅依赖ε而不依赖t_0时,则称x_e是一致稳定和一致渐近稳定的.

5. 当1,2和4中的$\delta(\varepsilon,t_0)$或$\delta(\varepsilon)$可以任意大时,则称x_e是全局渐近稳定或全局一致渐近稳定的.

李亚普诺夫稳定性是由李亚普诺夫(Ляпунов, A. M.)于1892年提出的.

李亚普诺夫函数(Liapunov function)　判定自由非线性系统原点稳定性的具有特殊符号性质的标量函数.令Ω是R^n原点$x=0$的某个邻域,$v(x,t)$为定义在$\Omega\times[0,+\infty)$上的标量函数,$v(x,t)$的符号性质指的是:

1. $v(x,t)$称为正定的(负定的),如果存在一正定(负定)函数$w(x)$,使得在Ω中有

$$v(x,t)\geqslant w(x)\ \ (v(t,x)\leqslant w(x)),$$

这里$w(x)$的正定(负定)性是指,$w(x)>0(w(x)<0)$,$\forall x(\neq0)\in\Omega$,且$w(0)=0$.

2. $v(x,t)$称为渐减的,如果存在正定函数$w(x)$,使得在Ω中有

$$|v(x,t)|\leqslant w(x),\ \forall t\geqslant t_0(t_0>0).$$

3. $v(x,t)$称为全局正定的、全局渐减的,如果性质1,2中的Ω为整个欧氏空间R^n.

利用函数$v(x,t)$的特殊符号性质,"李亚普诺夫稳定性"条目中自由非线性系统(1)在平衡状态$x_e=0$的稳定性有下列结果:

1. 如果在Ω中存在标量函数$v(x,t)$满足条件:$v(x,t)$正定,且沿"李亚普诺夫稳定性"条目中(1)的解$v(x,t)$的全导数$\dot v(x,t)|_{(1)}\leqslant0$对一切$t\geqslant t_0$,$x\in\Omega$成立,则(1)的原点$x=0$稳定,这里$\dot v(x,t)|_{(1)}$的含义如下:

$$\dot v(x(t),t)\Big|_{(1)}\triangleq\frac{\partial v(x,t)}{\partial t}+\frac{\partial v(x,t)}{\partial x}f(x,t)\Big|_{x=x(t)}.$$

2. 如果1中的$v(t,x)$还满足条件:$v(x,t)$渐减,即存在正定函数$w(x)$,使得在Ω内有

$$|v(x,t)|\leqslant w(x),\ \forall t\geqslant t_0,$$

则(1)的原点$x=0$一致稳定.

3. 如果在Ω内存在标量函数$v(x,t)$满足:$v(x,t)$正定,$\dot v(x,t)|_{(1)}$负定,则(1)的原点$x=0$渐近稳定.

4. 如果在Ω内存在标量函数$v(x,t)$满足:$v(x,t)$正定、渐减,且$\dot v(x,t)|_{(1)}$负定,则(1)的原点$x=0$一致渐近稳定.

5. 如果在R^n内存在标量函数$v(x,t)$满足:$v(x,t)$全局正定,$\dot v(x,t)|_{(1)}$全局负定,则(1)的原点$x=0$全局渐近稳定.

6. 如果在R^n内存在标量函数$v(x,t)$满足:$v(x,t)$全局正定、渐减,且$\dot v(x,t)|_{(1)}$全局负定,则(1)的原点$x=0$全局一致渐近稳定.

上列诸结果中的标量函数$v(x,t)$皆具有特殊符号性质,通称为李亚普诺夫函数.

自由定常非线性系统稳定性若干结果(several results of stability for free time-invariant nonlinear system)　自由定常非线性系统稳定性的几个主要定理.不受控系统

$$\dot x=f(x),\ f(0)=0 \tag{1}$$

称为自由定常非线性系统,其中$f(x)$是定义在原点$x=0$邻域$\Omega(\subset R^n)$的非线性n维向量函数,对(1)中f的要求及有关记号与"李亚普诺夫函数"条目中的相应要求相同.对于自由定常非线性系统原点稳定性的结果,除在"李亚普诺夫函数"条目中通过用$v(x)$代替$v(x,t)$而得到的外,还有:

1. 如果在Ω内,存在标量函数$v(x)$,它沿(1)的解的全导数$\dot v(x)|_{(1)}$为正定(负定)的,而$v(x)$本身不是半负定(半正定)的,则(1)的原点$x=0$不稳定.

2. 如果在Ω内,存在标量函数$v(x)$,使

$$\dot v(x)|_{(1)}=\lambda v(x)+w(x),$$

其中$\lambda>0$,$w(x)\geqslant0$,而$v(x)$不是半负定的,则(1)的原点不稳定.

3. 如果在Ω内,存在正定函数$v(x)$,使得

$$\dot v(x)|_{(2)}=w(x)\leqslant0,$$

且由$w(x)=0$所确定的点集$M=\{x|w(x)=0\}$除原点$x=0$外,不再包含(1)的其他整轨线,则(1)的原点渐近稳定.

4. 如果存在无限大正定函数,即对任意正数

N，都有正常数 r，使得在球：$\|x\|^2=r^2$ 的外部，有 $v(x)>N$，且 $\dot{v}(x)|_{(1)}=w(x)$ 是负半定的，而在集合 $M=\{x|w(x)=0\}$ 上除原点外不包含(1)的整轨线，则(1)的原点是全局渐近稳定的.

5. 如果存在正定矩阵 B，使对称矩阵 $BJ+J^TB$ 的特征值 $\lambda_i=\lambda_i(x)$，$i=1,2,\cdots,n$ 满足不等式：

$$\lambda_i<-\delta\quad(i=1,2,\cdots,n),$$

其中 δ 为某一正数，$J=J(x)=\frac{\partial f(x)}{\partial x}$，则系统(1)的原点渐近稳定.

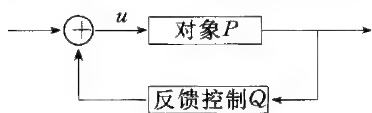
镇定(stabilization)　一个控制系统设计问题. 通过设计状态(或输出)反馈或其他补偿手段以得到系统稳定的设计称为镇定. 对于非线性系统，由于存在多种稳定性定义，例如李亚普诺夫稳定、有界输入及有界输出稳定等，它们之间存在联系但并不等价，从而存在多种镇定问题. 镇定研究的重点在于线性部分出现不能控或不能观的虚轴上的极点的情形. 研究的方法有经典的李亚普诺夫方法、中心流形方法、松弛反馈和算子分解等方法. 一般的反馈镇定问题还没有解决.

最小相位非线性系统(minimal phase nonlinear system)　其零动态为李亚普诺夫意义下渐近稳定的一类非线性控制系统. 最小相位原是线性控制系统理论中的概念，由本斯(Byrnes, C.)、依斯朵里(Isidori, A.)等引入到非线性控制系统研究中. 记

$$\dot{x}_2=f(0,x_2,u^*)$$

是一般非线性系统的零动态(参见"零动态")，如果它是李亚普诺夫意义下的渐近稳定系统，那么称整个系统是最小相位的，否则称非最小相位的.

反馈控制(feedback control)　控制策略的一种特殊方式. 反馈控制是控制理论和控制工程中最重要的一种控制策略. 它是用对象的状态或者输出来改造控制输入，使得系统有期望的性能(上图). 图中的 Q 是非线性算子时，称为非线性反馈控制. 常用的反馈控制有动态反馈、状态反馈和输出反馈三种.



微分几何方法(differential geometric approach)　研究非线性控制系统的一种方法. 利用现代微分几何框架来研究非线性系统的性能和反馈律设计称为非线性控制系统理论的微分几何方法. 主要应用的是微分流形、李群、李代数等理论，讨论对象多为仿射非线性系统.

非线性控制系统理论的微分几何方法在 20 世纪 70 年代提出，80 年代获得了很大的发展，逐渐形成了一套新的研究非线性控制系统的方法. 微分几何方法能够将线性系统几何方法的许多结论平行地推广到仿射非线性系统. 目前应用微分几何方法对于线性化、干扰解耦、输入-输出解耦和模型匹配等课题都获得了很好的成果.

分解方法(decomposition approach)　研究控制的分析和设计的一种方法. 对于非线性控制系统，其分解方法的来由可以追溯到线性控制系统的多项式方法. 1984 年，汉默(Hammer, J.)最早讨论了离散非线性系统的右互质分解，至今分解方法主要用于镇定和鲁棒分析领域. 记 $G:u\to v$ 为一个非线性系统，U 和 V 是线性空间，U_s 和 V_s 分别是 U 和 V 的稳定子空间，如果 $GU_s\subset V_s$，则 G 称为稳定的. 记 $D_0(G)=\{u\in U_s, Gu\in V_s\}$. 如果存在两个稳定系统 N 和 D，其中 $N:U\to V$，$D:U\to U$ 是可逆的，使得 $G=ND^{-1}$，那么称 G 存在右分解. 如果进一步有 $D_0(G)=DU_s$，那么称 G 存在右互质分解. 类似地可定义 G 的左分解和左互质分解. 对于非线性算子 G，它的右互质分解不总是存在的，从而非线性控制系统分解方法中一个大问题就是研究分解存在的条件，对离散非线性系统，这个条件已经找到.

非线性调节(nonlinear regulation)　设计反馈控制规律使相应的闭环系统成为李亚普诺夫意义下渐近稳定的过程. 对于形如下的非线性控制系统

$$\begin{cases}\dot{x}=f(x,u),\\ y=h(x),\end{cases}$$

如果存在依赖于状态 $x(\in\mathbb{R}^n)$ 或依赖于量测输出 $y(\in\mathbb{R}^n)$ 的控制向量值函数 $u(x)$ 或 $u(y)$，使得将 $u(x)$ 或 $u(y)$ 作用于非线性系统而形成的闭环系统

$$\dot{x}=f(x,u(x))$$

或

$$\dot{x}=f(x,u(h(x)))$$

是李亚普诺夫意义下渐近稳定的系统，则说非线性系统是能调节的；如果闭环系统是全局渐近稳定的，则说非线性系统是能全局调节的；如果闭环系统在包含原点 $x=0$ 为内点的某区域 D 内是渐近稳定的，则说非线性系统在 D 内是能调节的；如果闭环系统在原点 $x=0$ 的某邻域内是渐近稳定的，则说非线性系统是局部能调节的. 使闭环系统成为渐近稳定的控制函数 $u(x)$ 或 $u(y)$ 称为非线性系统的(静态)状态调节器或量测输出调节器. 当调节器 $u(x)$ 是 x 的线性向量值函数，即 $u(x)=Kx(K:r\times n$ 矩阵)时，称 $u(x)$ 是非线性系统的线性状态调节器，并说非线性系统能通过线性状态反馈进行调节；当 $u(x)$ 是 x 的非线性向量值函数时，称 $u(x)$ 是非线性系统的非线性状态调节器，并说非线性系统能通过非线性状态反馈进行调节. 同样可有线性和非线性输出反馈调节器.

非线性 H_∞ 控制(nonlinear H_∞ control)　一种抑制外部干扰的非线性系统控制方法. 这种新方法

是20世纪90年代兴起的.考虑如下系统:

$$\begin{cases}\dot{x}=f(x,u,w),\\ y=h_1(x,u,w),\\ z=h_2(x,u,w),\end{cases}$$

其中x是状态变量,y是量测输出变量,z是调节输出变量,w是外部信号(可以是干扰信号),u是控制变量,f是非线性向量场,h_1,h_2皆为非线性向量函数.所谓非线性H_∞控制问题乃是要求找到(设计)控制函数$u(\cdot)$(可以是状态x的函数,也可以是量测输出y的函数),使得在$u(\cdot)$的作用下有:

1. 当$w=0$时,闭环系统$\dot{x}=f(x,u(\cdot),0)$渐近稳定.

2. 存在$\gamma>0$,使得当$x(0)=0$时,L_2增益不等式

$$\int_0^t z^{\mathrm{T}}(t)z(x)\mathrm{d}x\leqslant\gamma^2\int_0^t w^{\mathrm{T}}(x)w(x)\mathrm{d}x$$

成立,$\forall t>0$.

非线性H_∞控制源于20世纪80年代发展起来的线性H_∞控制理论.鉴于非线性系统的多样性和复杂性,线性H_∞控制的许多分析方法不易直接利用.目前主要的研究方法是利用微分对策的观点,将非线性H_∞控制问题化为HJI(Hamilton-Jacobi-Isaacs)方程(或不等式)的求解问题.由于对一般非线性系统,HJI方程(不等式)的求解甚至解的存在性问题都不易解决,至今只对一些特殊的非线性系统得到较为具体的设计方法.非线性H_∞控制有广泛的应用背景.

齐次系统(homogeneous system) 其数学描述具有齐次特性的非线性系统.构成这类系统的向量场$f(x)$具有齐次性质,即

$$f(\varepsilon x)=\varepsilon^k f(x),\ x\in\mathrm{R}^n,$$

其中ε为实数,k为自然数.另一类更广的齐次性质乃是:存在一个扩张(dilation)向量

$$r=(r_1,r_2,\cdots,r_n)\in\mathrm{R}^n, r_i>0\ (i=1,2,\cdots,n)$$

和实数$\varepsilon>0$,使得

$$f_i(\varepsilon^{r_1}x_1,\varepsilon^{r_2}x_2,\cdots,\varepsilon^{r_n}x_n)=\varepsilon^{k+r_i}f_i(x_1,x_2,\cdots,x_n),$$

其中$f(x)=[f_1(x),f_2(x),\cdots,f_n(t)]^{\mathrm{T}}$.齐次系统有许多优良性质,研究它有助于对一般非线性系统的了解.

格拉斯曼流形方法(Grassman manifold method) 通过格拉斯曼流形来描述系统的一种方法.将n维空间中的一个m维子空间作为一个点,这种等价关系作成的子流形,称为一个格拉斯曼流形,记为G_m^n.对于线性系统

$$\dot{x}=Ax+Bu,\ x\in\mathrm{R}^n,\ u\in\mathrm{R}^m,$$

$\mathrm{Span}\{B\}$可看成G_m^n中的一个点.e^{At}是G_m^n上的一个群作用.因此,系统能控就等价于从B出发的群轨道跑遍G_m^n.这种方法不仅对分析线性系统有许多方便之处,对于非线性系统也有许多应用.

哈密顿控制系统(Hamilton control system) 一种基于哈密顿向量场的控制系统.设H为系统的哈密顿函数,$G_i\ (i=1,2,\cdots,m)$为输入哈密顿函数,则系统可表示为

$$\begin{cases}\dot{x}=X_H+\sum\limits_{i=1}^{m}X_{G_i}u_i,\\ y=\{H,G_i\},\end{cases}$$

这里$X_H(X_{G_i})$为由$H(G_i)$生成的哈密顿向量场,$\{\cdot,\cdot\}$为泊松括号.哈密顿系统有广泛的物理背景及工程背景.

变结构控制

变结构控制(variable structure control) 控制系统的一种非线性综合方法.它的基本原理是:当系统状态变量x的某个函数$s(x)$大于零时,控制u取为$u=u^+(x)$,而当$s(x)$小于零时,控制u则取为$u=u^-(x)$,$u^-(x)\neq u^+(x)$.这里$s(x)$称为切换函数.在变结构控制系统中,由于描述系统运动的微分方程的右端函数在切换面$\varphi\triangleq\{x\in\mathrm{R}^n:s(x)=0\}$上是不连续的,所以变结构控制是一种不连续控制,这里R^n表示系统的状态空间.在变结构控制系统中,无论系统的初始状态如何,系统将在有限时间内到达并维持在切换面上运动,且逐渐向原点趋近,这种运动称为滑动模态.滑动模态是变结构控制系统的特征.因此变结构控制又称为滑模控制.变结构控制系统的突出优点是其滑动模态对一类干扰及参数摄动具有不变性(完全自适应性),其突出缺点是可能出现的抖振.变结构控制系统的设计包括以下两方面:

1. 设计切换函数,使得相应的滑动模态具有所期望的品质,如渐近稳定、二次性能指标最优等.

2. 设计变结构控制器,使得系统自任意初始状态出发,总能在有限时间内到达并维持在切换面上运动.

变结构控制目前已被用于系统的镇定、运动跟踪、模型跟踪、自适应控制、状态观测等方面,并被推广到离散系统、非线性系统、大系统、时滞系统、分布参数系统、不确定性系统、随机系统、模糊系统等系统中去.在实际系统中也得到广泛应用,如飞机控制、卫星姿态控制、机器人控制、电机控制、电力系统控制、柔性空间结构控制等.

变结构控制系统中微分方程的定义(definition of differential equation in variable structure control system) 变结构控制系统中不连续微分方程定义的确定方法.在变结构控制系统中,描述系统运动的微分方程的右端函数在切换面上是不连续的,因而

不能用经典的微分方程理论来描述系统的运动. 处理这一问题的一般方法有:

1. 广义动态系统理论.

2. 非力波夫理论,见"非力波夫理论".

3. 化为相坐标方法,如果一变结构控制系统具有如下的相坐标表示:

$$\dot{x}_1 = x_2,$$
$$\dot{x}_2 = x_3,$$
$$\cdots\cdots$$
$$\dot{x}_{n-1} = x_n,$$
$$\dot{x}_n = \alpha(x) + u_0,\ x = [x_1, x_2, \cdots, x_n]^{\mathrm{T}},$$
$$\alpha(x) = \begin{cases} \alpha^+(x),\ s(x) > 0, \\ \alpha^-(x),\ s(x) < 0, \end{cases}$$

则系统在切换面上的运动的微分方程为

$$s(x_1, x_2, \cdots, x_{n-1}, x_n) = s(x_1, \dot{x}_1, \cdots, x_1^{(n-2)}, x_1^{(n-1)}) = 0.$$

4. 等价控制方法(参见"等价控制").

上述方法 1 和 2 适用于一般的右端不连续微分方程,而方法 3 和 4 则只适用于具有滑动模态的变结构控制系统.

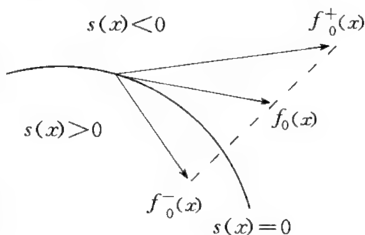
非力波夫理论(Filippov theory) 处理右端不连续微分方程的一种理论. 一个微分方程

$$\dot{x} = f(t, x)$$

的非力波夫解 $x(t)$ 定义为:$x(t)$ 绝对连续且满足

$$\dot{x}(t) \in \bigcap_{\delta>0}\bigcap_{\mu N=0} \overline{\omega x} f(t, \mathscr{B}(x,\delta)\backslash N)\ (\mathrm{a.e.}),$$

这里 $\mathscr{B}(x,\delta)$ 表示 R^n 中以点 x 为球心、以 δ 为半径的开球,μN 表示 R^n 中集合 N 的勒贝格测度,ωx 及 $\overline{\omega x}$ 分别表示凸组合及其闭包. 在变结构系统中,向量场 f 在切换面上的取值一般来说是不确定的,上述定义中的零测集 N 的作用就是用来消除这种不确定性. 当切换面的维数为 $n-1$ 维时,变结构系统在切换面上的非力波夫解的几何意义是:它所确定的向量场 f_0 与切换面相切,且其末端落在两向量场向量 f_0^+,f_0^- 末端的连线上,如上图所示. 这里 f_0^+,f_0^- 分别为系统的两种结构所确定的在切换面附近的极限向量场.



等价控制(equivalent control) 使变结构控制系统保持在其切换面上运动的控制. 当变结构控制系统存在滑动模态时,可以用等价控制来消除其微分方程右端在切换面上的不确定性. 对变结构控制系统

$$\dot{x} = f(x,t) + B(x,t)u,$$

$$u = \begin{cases} u^+(x,t),\quad 当\ s(x) > 0, \\ u^-(x,t),\quad 当\ s(x) < 0, \end{cases}$$

令 $G = \partial s/\partial x$,当 GB 非奇异时,在切换面上补充定义控制 u 的值,它使得沿着系统的解,恒有 $\mathrm{d}s/\mathrm{d}t \equiv 0$,由此所得到的控制就称为等价控制,一般记为 u_{eq}. 对于上例,

$$u_{\mathrm{eq}} = -[G(x)B(x,t)]^{-1}G(x)f(x,t),$$

系统在切换面上的运动按如下方程定义:

$$\begin{cases} \dot{x} = f(x,t) + B(x,t)u_{\mathrm{eq}}(x,t), \\ s(x) = 0. \end{cases}$$

等价控制的含义是:一旦系统到达切换面,等价控制将使之维持在切换面上. 等价控制的一个严重缺点是 u_{eq} 要求严格实现,因此不具鲁棒性.

切换面(switching surface) 变结构控制系统的切换函数等于零所确定的曲面. 设 $s(x)$ 为一变结构控制系统的切换函数,则切换面定义为

$$\varphi = \{x \in \mathrm{R}^n : s(x) = 0\},$$

此处 R^n 为系统的状态空间,当 s 为 x 的线性函数时,切换面是 R^n 中的一个超平面. 若 s 是 m 维向量,则切换面共有 m 个.

滑动模态(sliding mode) 变结构控制系统中发生并维持在切换面上的运动形式. 考虑如下系统

$$\dot{x} = f(x,t),\ x \in \mathrm{R}^n,$$

对于 R^n 中的一个子集 M,如果 $\forall x(t_0) \in M$,总有 $x(t) \in M, t \geqslant t_0$,则称 M 为系统的不变集,又称系统在 M 上的运动为不变运动. 变结构控制系统在切换面上的不变运动称为滑动模态(滑动运动). 设多变量变结构控制系统的切换函数为

$$s(x) = [s_1(x), s_2(x), \cdots, s_m(x)]^{\mathrm{T}},$$

令 $\varphi_i = \{x \in \mathrm{R}^n : s_i(x) = 0\}$,滑动模态的可能类型有:

1. 在 $\varphi_i (i = 1, 2, \cdots, m)$ 上的滑动运动.

2. 在 $\varphi_i \cap \varphi_j (i, j = 1, 2, \cdots, m, i \neq j)$ 上的滑动运动.

……………………………………………

m. 在 $\varphi_0 = \varphi_1 \cap \varphi_2 \cap \cdots \cap \varphi_m$ 上的滑动运动.

但有些研究者仅定义 φ_0 上的运动为滑动模态.

线性系统切换函数的设计(design of switching function for linear system) 设计线性系统变结构控制的切换函数的方法. 考虑线性控制系统

$$\dot{x} = Ax + Bu,\ x \in \mathrm{R}^n,\ u \in \mathrm{R}^m \tag{1}$$

的变结构控制问题. 其切换函数设计的主要方法有:

1. 变换为简约型. 通过坐标变换将(1)化为如下形式:

$$\begin{cases} \dot{x}_1 = A_{11}x_1 + A_{12}x_2, \\ \dot{x}_2 = A_{21}x_1 + A_{22}x_2 + B_2 u, \end{cases} \tag{2}$$

取切换函数 $s = C_1x_1 + C_2x_2$,在 C_2 可逆的情况下,系统(2)滑动模态运动方程为

$$\begin{cases}\dot{x}_1 = (A_{11} - A_{12}C_2^{-1}C_1)x_1,\\ \dot{x}_2 = -C_2^{-1}C_1x_1.\end{cases}$$

可以证明，当(A,B)能控时，(A_{11},A_{12})亦能控，因而滑动模态的动力学部分可以通过线性系统理论中极点配置、二次性能指标最优等方法解决，从而可确定切换函数中C_1和C_2的值.

2. 运动分解方法. 对系统(2)作状态变换

$$\begin{cases}\dot{x}_1 = z,\\ \dot{x}_2 = C_2^{-1}s - C_2^{-1}C_1z,\end{cases}$$

则系统(2)化为

$$\dot{z}=(A_{11}-A_{12}C_2^{-1}C_1)z+A_{12}C_2^{-1}s,$$
$$\dot{s}=[(C_1A_{11}+C_2A_{21})-(C_1A_{12}+C_2A_{22})C_2^{-1}C_1)]z+(C_1A_{12}+C_2A_{22})C_2^{-1}s+C_2B_2u.$$

于是，系统的运动可以看做由两部分变量来表征：其一是由上述滑动运动的状态变量z，其二是描述到达过程的状态变量s. 系统滑动运动的微分方程为

$$\dot{z} = (A_{11} - A_{12}C_2^{-1}C_1)z,$$

其稳定性或其他性能指标可用线性系统的经典方法来解决.

非线性系统切换函数的设计(design of switching function for nonlinear system) 设计仿射非线性系统变结构控制的切换函数的方法. 考虑仿射非线性控制系统

$$\dot{x} = f(x) + g(x)u,\ x \in \mathbb{R}^n,\ u \in \mathbb{R}^m, \quad (1)$$

其切换函数的设计主要有以下几种方法：

1. 简约型. 把系统(1)变换为如下形式：

$$\begin{cases}\dot{x}_1 = A_1(x) = A_1(x_1,x_2),\\ \dot{x}_2 = A_2(x) + B_2(x)u,\end{cases} \quad (2)$$

其中$x_1\in\mathbb{R}^{n-m}$，$x_2\in\mathbb{R}^m$，$x=[x_1^{\mathrm{T}},x_2^{\mathrm{T}}]^{\mathrm{T}}$，系统(2)称为简约型. 如果存在函数$\omega(x_1)\in\mathbb{R}^m$，使得子系统

$$\dot{x}_1 = A_1(x_1,\omega(x_1))$$

渐近稳定，这时系统(2)的切换函数可取为

$$s(x) = x_2 - \omega(x_1).$$

2. 用相变量表示的正则型. 如果系统(1)可化为如下形式：

$$\begin{cases}\dot{x}_{i,j}=x_{i,j+1}\ (i=1,2,\cdots,m;j=1,2,\cdots,n_i-1),\\ \dot{x}_{i,n_i}=\alpha_i(x)+\beta_i(x)u_i\ \ (i=1,2,\cdots,m),\end{cases} \quad (3)$$

其中$\sum\limits_{i=1}^{m}n_i = n$，系统(3)称为用相变量表示的正则型，此时系统(3)的切换函数可取为

$$s_i(x) = C_{i,1}x_{i,1} + \cdots + C_{i,n_i-1}x_{i,n_i-1} + x_{i,n_i}\quad (i = 1,2,\cdots,m),$$

其中$\{C_{i,1},\cdots,C_{i,n_i-1}\}$使得

$$\lambda^{n_i-1} + C_{i,n_i-1}\lambda^{n_i-2} + \cdots + C_{i,2}\lambda + C_{i,1}$$

为稳定多项式.

3. 正则型. 若系统(1)能化为如下形式：

$$\begin{cases}\dot{z}_{i,j} = z_{i,j+1}\quad (i = 1,2,\cdots,m;\\ \qquad j = 1,2,\cdots,r_i - 1),\\ \dot{z}_{i,r_i} =\alpha_i(z,\eta) + \sum\limits_{k=1}^{m}\beta_{ik}(z,\eta)u_k\\ \qquad (i = 1,2,\cdots,m),\\ \dot{\eta} = \gamma(z,\eta),\end{cases} \quad (4)$$

其中$z=[z_{1,1},\cdots,z_{1,r_1};\cdots;z_{m,1},\cdots,z_{m,r_m}]^{\mathrm{T}}\in\mathbb{R}^r$，$r=\sum\limits_{i=1}^{m}r_i$，$\eta\in\mathbb{R}^{n-r}$，系统(4)称为正则型. 如果$\dot{\eta}=\gamma(0,\eta)$是渐近稳定的，则系统(4)的切换函数可取为

$$s_i=C_{i,1}z_{i,1}+\cdots C_{i,r_i-1}z_{i,r_i-1}+z_{i,r_i}\ (i=1,2,\cdots,m),$$

其中$\{C_{i,1},\cdots,C_{i,r_i-1}\}$，使得

$$\lambda^{r_i-1} + C_{i,r_i-1}\lambda^{r_i-2} + \cdots + C_{i,2}\lambda + C_{i,1}$$

为稳定多项式. 对于一般的非线性系统，切换函数的设计问题目前仍然没有解决.

滑动模态对干扰的不变性(invariance of sliding mode on disturbance) 滑动模态的抗干扰性质. 对于如下受扰线性系统

$$\dot{x}= (A + \Delta A)x + (B + \Delta B)u + d(t),$$
$$s = Cx,$$

当系统的参数摄动ΔA，ΔB及外部干扰$d(t)$满足如下条件：存在$\widetilde{\Delta A}$，$\widetilde{\Delta B}$及$\widetilde{d(t)}$，使得

$$\begin{cases}\Delta A = B\cdot\widetilde{\Delta A},\\ \Delta B = B\cdot\widetilde{\Delta B},\\ d(t) = B\cdot\widetilde{d(t)}\end{cases} \quad (1)$$

时，滑动模态的运动方程将与ΔA，ΔB及$d(t)$无关，变结构控制系统的这一性质称为滑动模态对干扰的不变性或完全自适应性，条件(1)称为匹配条件. 当匹配条件成立时，虽然ΔA，ΔB及$d(t)$不影响滑动模态，但却影响到达过程，当ΔA，ΔB及$d(t)$有界且界已知时，可以适当设计变结构控制器，使得到达过程对ΔA，ΔB及$d(t)$也是鲁棒的，因而整个变结构控制系统对ΔA，ΔB及$d(t)$是鲁棒的. 上述性质对时变系统及非线性系统均成立. 当干扰或参数摄动不满足匹配条件时，对到达过程及滑动模态的研究都是复杂的.

到达条件(arrival condition) 变结构控制中保证系统状态在有限时间内运动到滑动模态的条件. 直到20世纪80年代末期，变结构控制器主要是通过如下的到达条件来设计的：

1. 设计$u=u^{\pm}(x)$，使得当$s_i<0$，$\dot{s}_i>0$，$\lim\limits_{s_i\to 0^+}\dot{s}_i<0$；当$s_i>0$时，$\dot{s}_i<0$，$\lim\limits_{s_i\to 0^-}\dot{s}_i>0$.

2. 令$V=\frac{1}{2}s^{\mathrm{T}}s$. 设计$u=u^{\pm}(x)$，使得

$$\dot{V} = s^{\mathrm{T}}\dot{s} < 0,\ 且\ \lim_{\|s\|\to 0}\dot{V}< 0.$$

对单变量控制系统，通过上述条件设计变结构控制器是不难的. 但对多变量控制系统，由于上述条件为 $2m$ 个(m 为 s 的维数)条件不等式，由此来设计 $u^{\pm}(x)$是非常困难的，克服这一困难的方法是采用趋近律方法.

趋近律(reaching law)　一种等式形式的到达条件. 一般的趋近律为

$$\dot{s} = -\varepsilon \operatorname{sgn} s - f(s),$$

其中

$$\varepsilon = \operatorname{diag}\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m\} \quad (\varepsilon_i > 0),$$

$$\operatorname{sgn} s = [\operatorname{sgn} s_1, \operatorname{sgn} s_2, \cdots, \operatorname{sgn} s_m]^{\mathrm{T}},$$

$f(x) \in \mathrm{R}^m$ 满足 $f(0)=0$；当 $s \neq 0$ 时，$s^{\mathrm{T}} f(s) > 0$.

选取不同的 ε 及 f，可以得到各种各样的趋近律，如：

1. 等速趋近律：$\dot{s} = -\varepsilon \operatorname{sgn} s$.

2. 指数趋近律：

$$\dot{s} = -\varepsilon \operatorname{sgn} s - ks,$$

$$k = \operatorname{diag}\{k_1, k_2, \cdots, k_m\} \ (k_i > 0).$$

3. 幂次趋近律：

$$\dot{s}_i = -k_i |s_i|^{\alpha} \operatorname{sgn} s_i, \ k_i > 0, \ 0 < \alpha < 1.$$

通过趋近律，求解变结构控制器问题便从以往的求解 $2m$ 个条件不等式问题简化为求解 m 个代数方程问题，这大大方便了控制器的设计. 这一点对于一些复杂的变结构控制系统，如非线性系统、时滞系统等显得尤为重要. 变结构控制系统在达到滑动模态之前的运动称为到达过程. 通过选择趋近律中的参数可以保证到达过程的品质并减弱变结构系统中的抖振现象.

切换模式(switching pattern)　系统到达滑动模态的方式. 它是多变量变结构系统中出现的一个特殊问题. 令各切换面

$$\varphi_i = \{x \in \mathrm{R}^n : s_i(x) = 0\} \ (i = 1, 2, \cdots, m; m \geqslant 2),$$

则系统的滑动面 φ 为各 $\varphi_i (i=1,2,\cdots,m)$之交，即 $\varphi = \varphi_1 \cap \varphi_2 \cap \cdots \cap \varphi_m$. 常用的切换模式有：

1. 固定递阶切换模式. 这种切换模式按照预先规定的顺序把系统依次驱动到相应的切换面上，例如按如下顺序：

$$x(t_0) \to \varphi_1 \to \varphi_1 \cap \varphi_2 \to \cdots \to \varphi_1 \cap \varphi_2 \cap \cdots \cap \varphi_m = \varphi.$$

即使系统的状态在到达 φ_1 之前先到达 φ_2，也仍然有可能离开 φ_2 而趋向 φ_1，并在 φ_1 上滑动，而后才在 $\varphi_1 \cap \varphi_2$上滑动.

2. 自由切换模式. 这种切换模式同时启动所有的控制变量，使系统向各个切换面运动，到达后就在该切换面上滑动，直到最后到达并维持在 φ 上运动. 至于系统首先到达哪个切换面并在其上滑动，完全由系统运动的自然趋势所决定.

3. 最终滑动模态切换模式. 这种切换模式只要求系统的状态最终到达滑动面 φ，至于在各个切换面 $\varphi_1, \varphi_2, \cdots, \varphi_m$ 上是否有滑动运动，并不关心.

固定递阶切换模式把多变量控制器的设计问题化为单变量来处理，但却人为地增长了到达过程的时间. 自由切换模式及最终滑动模态切换模式克服了这一缺点，前者可用趋近律来设计，后者可用单位向量控制来实现.

分散切换模式(decentralized switching pattern)　系统到达滑动模态的方式. 是将系统分解成若干子系统，各子系统到达滑动模态的方式.

线性系统变结构控制的设计(design of variable structure control for linear system)　线性定常系统变结构控制器的设计方法和步骤. 考虑如下线性系统 $\dot{x} = Ax + Bu$，其变结构控制器的设计可按如下步骤进行：

1. 运动分解. 将上述系统变换为

$$\dot{x}_1 = A_1 x,$$

$$\dot{x}_2 = A_2 x + B_2 u,$$

式中 $x = [x_1^{\mathrm{T}}, x_2^{\mathrm{T}}]^{\mathrm{T}}$.

2. 确定切换函数. $s = Cx = C_1 x_1 + C_2 x_2$，使得相应的滑动模态渐近稳定(参见“线性系统切换函数的设计”).

3. 利用到达条件设计变结构控制律，例如可用趋近律求解变结构控制器. 令 $\dot{s} = -\varepsilon \operatorname{sgn} s - f(s)$，其中 $f(s)$满足 $s f(s) > 0$ 的任一连续函数. 将上式与

$$\dot{s} = C\dot{x} = C_1 A_1 x + C_2 A_2 x + C_2 B_2 u$$

联立，得变结构控制器

$$u = -(C_2 B_2)^{-1} [C_1 A_1 x + C_2 A_2 x + \varepsilon \operatorname{sgn} s + f(s)],$$

其中 C 的设计应保证 $C_2 B_2$ 非奇异.

非线性系统变结构控制的设计(design of variable structure control for nonlinear system)　非线性系统变结构控制器的设计方法和步骤. 考虑如下仿射非线性系统 $\dot{x} = A(x) + B(x)u$. 其变结构控制器的设计可按如下步骤进行：

1. 运动分解. 将上述系统变换为

$$\dot{x}_1 = A_1(x),$$

$$\dot{x}_2 = A_2(x) + B_2(x)u.$$

2. 确定切换函数 $s = C(x)$，使得相应的滑动模态渐近稳定(参见“非线性系统切换函数的设计”).

3. 利用到达条件设计变结构控制律，例如可用趋近律求解变结构控制器. 令

$$\dot{s} = -\varepsilon \operatorname{sgn} s - f(s),$$

其中 $f(s)$是满足 $s f(s) > 0$ 的任一连续函数. 将上式与

$$\dot{s} = \frac{\mathrm{d}}{\mathrm{d}t} C(x)$$

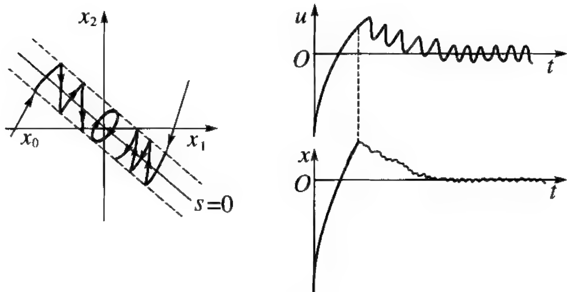
$$= \frac{\partial C}{\partial x_1}A_1(x) + \frac{\partial C}{\partial x_2}A_2(x) + \frac{\partial C}{\partial x_2}B_2(x)u$$

联立,得变结构控制器

$$u = -\left(\frac{\partial C}{\partial x_2}B_2(x)\right)^{-1}\left(\frac{\partial C}{\partial x_1}A_1(x) + \frac{\partial C}{\partial x_2}A_2(x) + \varepsilon \operatorname{sgn} s + f(s)\right),$$

其中 C 的设计应保证 $\frac{\partial C}{\partial x_2}B_2(x)$ 可逆.

抖振(chattering) 变结构系统在其滑动面附近的高频振荡. 在变结构系统中,由于开关器件(如继电器)的时滞及回滞等特性,以及执行机构的时延及惯性等实际因素的影响,系统的状态到达切换面后,将不是保持在其上作滑动运动,而是在切换面附近作高频来回穿越运动,这种现象称为抖振. 其在相平面及时域中的表现如下图所示.



抖振现象是不连续控制系统中普遍存在的现象,它是一种高频振荡,有可能激励起实际系统中未建模的高频运动成分,这一般不是所期望的,因而削弱抖振是变结构系统的一个重要课题.

削弱抖振的基本方法是对控制进行连续化,如用饱和函数 sat$s$ 代替符号函数 sgn$s$,或减小控制器在切换面附近的跳变幅度. 调节趋近律中的参数也是减小抖振的一个有效措施.

离散系统变结构控制(variable structure control of discrete system) 对离散系统设计变结构控制的方法. 与连续系统比较类似. 在离散系统变结构控制中出现的一个特殊问题是所谓的准滑动模态. 由于在一个采样周期内,控制量的大小是不变的(采样保持),所以本质上,在采样时刻,系统状态恰好位于切换面上的概率为零,故除初始条件位于切换面上的运动外,系统的运动将在切换面附近作来回穿越运动,这两种运动统称为准滑动模态. 准滑动模态是离散变结构系统的固有特点,与连续变结构系统中的抖振现象是有区别的. 准滑动模态存在的区域称为准滑动模态带,其边界是平行于切换面的超曲面.

大系统的变结构控制(variable structure control for large scale system) 对大系统设计变结构控制的方法. 其关键思想是利用滑动模态对干扰的不变性原理. 这里把各子系统的关联项作为干扰对待,当关联矩阵满足匹配条件时,就可以设计各子系统的切换函数,它仅为该子系统状态变量的函数. 如此,相应的滑动模态则为若干个解耦子系统的滑动模态. 值得注意的是,一般情况下,变结构控制器仍和整个大系统的状态变量有关.

模型跟踪变结构控制(variable structure control for model tracking) 利用变结构控制对系统实现模型跟踪的方法. 跟踪问题是自动控制系统的基本任务之一,利用变结构控制实现模型跟踪是非常方便的. 其基本方法是:

1. 建立误差向量 $e(t) \triangleq x(t) - x_M(t)$ 的动力学模型,这里 $x(t)$,$x_M(t)$ 分别表示受控对象和参考模型的状态向量.

2. 构造关于 e 的切换函数 $s=Ce$,使得相应的滑动模态是渐近稳定的.

3. 设计变结构控制器实现该滑动模态.

运动跟踪、自适应控制均可利用此方法来实现. 利用变结构控制理论还可以实现模型到达,即使受控系统的状态在一定时间之后完全等同于参考模型的状态. 基本设计方法是构造一个 m 维(m 为控制变量的个数)动态补偿器,对整个扩充的动态系统,设计变结构控制器,使得相应的滑动模态与参考模型相吻合.

大规模系统

大规模系统(large scale system) 简称大系统. 一类系统. 是对一类规模大、维数高的系统的总称.

随着科学技术与社会生产水平的发展,规模庞大、结构复杂的系统日益增多. 同时计算机技术的飞速发展,使得控制和管理这些系统成为可能. 20 世纪 60 年代末,大规模系统理论及应用研究开始出现,自 70 年代起其发展特别迅速,逐渐形成一个专门的研究领域.

从大系统名称的直观含义知,大系统通常是指规模大、维数高的系统. 对于系统究竟多大才算是大系统,目前没有公认的定义. 在实际研究中,当系统的维数大到用常规的建模、分析的控制方法不能以合理的计算量来求解,而需要将系统分解为若干子系统来处理,或需将其控制器划分为几个部分,各自采用所能获取的部分信息来对系统进行控制时,就可认为该系统为大系统.

大系统理论是现代控制理论基础上发展起来的,现代控制理论中几乎各个方面的研究都已扩展到大系统理论研究中. 大系统理论研究的主要内容和成果有:

1. 大系统的建模及模型简化. 其中最主要的成果是大系统的模型降阶方法,包括时域中各种模型集结方法和频域中各种模型近似匹配降阶方法.

2. 大系统的稳定性分析. 其中最主要的成果是用向量李亚普诺夫方法与加权和李亚普诺夫方法导出的以各孤立子系统的稳定性和子系统间关联项刻画的各种大系统稳定性判据,以及 M 矩阵在简化这些判据中的应用.

3. 大系统的控制研究. 其中最主要的成果是关于利用局部信息进行大系统分散控制和固定模与分散控制关系的研究,及其以分解-协调方法为基础的大系统各种递阶优化算法的研究.

在应用方面,大系统理论在工农业生产、社会经济、资源管理、交通运输、环境保护等方面已有不少成功的应用实例.

大系统没有严格的数学定义. 文献中研究的大系统有如下特点:

1. 规模庞大. 通常系统中含有许多小系统,或跨越很大空间.

2. 结构复杂. 各小系统可能分级,控制采用的信息可分散,也可集中,分散信息又可分为完全分散或部分分散等.

3. 功能综合. 对一个大规模系统性能的评价是多指标、综合性的,大规模系统必须具备各种功能才能达到预期性能.

大小和复杂程度是个相对概念,随着科学技术的发展,对大规模系统的理解也可能发生变化.

互联系统(interconnected system) 亦称组合系统. 一种大系统. 是由许多小系统通过耦合项连结构成的大系统. 若系统的一般形式表示为

$$\frac{dx}{dt}=f(x,t),\quad x\in \mathsf{R}^m,\quad t\in \mathsf{R},$$

则该系统可分解成 n 个子系统,

$$\frac{dx_i}{dt}=f_i(x_i,t)+g_i(x,t)\quad (i=1,2,\cdots,n),$$

其中 $x=[x_1^{\mathrm{T}},x_2^{\mathrm{T}},\cdots,x_n^{\mathrm{T}}]^{\mathrm{T}}$, $f(x,t)=[f_1^{\mathrm{T}}(x_1,t)+g_1^{\mathrm{T}}(x,t),\cdots,f_n^{\mathrm{T}}(x_n,t)+g_n^{\mathrm{T}}(x,t)]^{\mathrm{T}}$, $g_i(x,t)$ 是耦合项,反映了其他子系统对第 i 个子系统的影响.

组合系统(composite system) 即“互联系统”.

大规模系统模型简化(model simplification for large scale system) 寻找一个简单的数学模型以代替原有的大规模系统模型的方法. 简化模型必须保留原系统的某些主要动态特性. 按系统用状态空间表示和传递函数表示,模型简化方法可分为时域法和频域法两类. 从工程观点看,有两个重要原因要简化模型:

1. 大系统计算量太大,以致系统仿真、分析和控制器设计都十分困难.

2. 根据简化模型设计的控制器,结构比较简单,容易工程实现.

集结法是模型简化的一种时域方法. 设系统 s_1 和 s_2 分别表示为

$$s_1:\ \dot{x}=Ax+Bu,$$
$$s_2:\ \dot{z}=Fz+Gu,$$

其中 A,B,F,G 是常数矩阵, $x\in \mathsf{R}^n$, $z\in \mathsf{R}^m$, $n>m$. 若存在常数矩阵 C,对一切 t 有 $z(t)=Cx(t)$,则称系统 s_2 是 s_1 的集结,C 称为集结矩阵. C 为集结矩阵的充分必要条件是 $FC=CA$, $G=CB$. 当 s_2 是 s_1 的集结时,F 的特征值集包含在 A 的特征值集中,通常将 A 的主导特征值包含在 F 中,因此简化系统 s_2 能反映 s_1 的主要动态特性.

大规模系统集结(aggregation for large scale system) 见“大规模系统模型简化”.

最优阶简化法(optimum method of reduced order) 一种模型集结降阶的最优化方法. 设原系统为

$$\begin{cases}\dot{x}=Ax+Bu,\\ y=Cx,\end{cases}$$

式中 $x\in \mathsf{R}^n$, $y\in \mathsf{R}^p$. 最优阶简化法是寻找降阶简化模型

$$\begin{cases}\dot{x}_r=A_rx_r+B_ru,\\ y_r=C_rx_r,\end{cases}$$

式中 $x_r\in \mathsf{R}^r$, $y_r\in \mathsf{R}^p$,使得输出误差

$$e(t)=y(t)-y_r(t)$$

的某个指标函数 J 极小化,通常取

$$J=\int_0^\infty e^{\mathrm{T}}(t)Qe(t)\mathrm{d}t,$$

其中 Q 是正定矩阵.

频域降阶法(frequency-domain method for reducing order) 大规模系统模型简化的一类频域方法. 主要有时间矩匹配、帕德近似、劳斯近似、连分式展开等.

时间矩匹配法(time-moment matching method) 大规模系统的一种频域降阶方法. 设 $G(s)$ 和 $h(t)$ 分别是系统的传递函数和脉冲响应函数,

$$G(s)=\frac{b_0+b_1s+\cdots+b_{n-1}s^{n-1}}{1+a_1s+\cdots+a_ns^n}.$$

称

$$\alpha_i=\int_0^\infty t^ih(t)\mathrm{d}t$$

为 $h(t)$ 的第 i 阶时间矩. 时间矩和传递函数系数满足关系式

$$\sum_{i=0}^{p}(-1)^i\frac{\alpha_i}{i!}a_{p-i}=b_p,\ a_0=1. \tag{1}$$

于是,已知 $G(s)$,即可求得各阶时间矩. 所谓时间矩匹配是要求依据原系统的 $2r$ 个时间矩 $\alpha_0,\alpha_1,\cdots,$

α_{2r-1}，去寻找一个低阶传递函数

$$G_r(s)=\frac{c_0+c_1s+\cdots+c_{r-1}s^{r-1}}{1+d_1s+\cdots+d_rs^r},$$

使其相对应的脉冲函数的前 $2r$ 个时间矩为 $\alpha_0,\alpha_1,\cdots,\alpha_{2r-1}$. 用时间矩匹配得到的简化模型并不保证和原系统有同样的稳定性质，目前在上述方法基础上提出了不少修改方案，以克服这个缺点.

帕德近似(Padé approximation)　有理函数逼近的一种方法. 可用于大规模系统在频域的降阶. 设 $G(s)$是系统的传递函数，对 $G(s)$进行幂级数展开得

$$G(s)=\sum_{i=0}^{\infty}c_is^i.$$

所谓帕德近似是要求寻找一个低阶传递函数

$$G_r(s)=\frac{d(s)}{e(s)}=\frac{d_0+d_1s+\cdots+d_ps^p}{1+e_1s+\cdots+e_rs^r}\quad(p\leqslant r),$$

使得

$$G(s)-G_r(s)=\sum_{j=p+r+1}^{\infty}w_js^j,$$

即

$$e(s)\sum_{i=0}^{\infty}c_is^i-d(s)=\sum_{j=p+r+1}^{\infty}u_js^j.$$

比较直到 $p+r$ 次幂的系数，得到关于 $G_r(s)$系数的线性代数方程，求解得到 $G_r(s)$. 帕德近似计算简单，对次数低于 $p+r$ 的多项式类型输入，简化模型和原系统输出相同. 但它不能保持原系统的稳定性，因此有不少修改方案，以克服这个缺点.

劳斯近似(Routh approximation)　一种基于递推计算的大规模系统的频域降阶法. 设 $G(s)$是系统的传递函数，

$$G(s)=\frac{b_1s^{n-1}+\cdots+b_n}{a_0s^n+\cdots+a_n},$$

$$\hat{G}(s)=\frac{1}{s}G\left(\frac{1}{s}\right)=\frac{b_ns^{n-1}+\cdots+b_0}{a_ns^n+\cdots+a_0}.$$

若 $\hat{G}(s)$是渐近稳定的，它有 α-β 规范展开，即

$$\hat{G}(s)=\sum_{k=1}^{n}\beta_k\prod_{j=1}^{k}F_j(s),$$

$$F_j(s)=\cfrac{1}{\alpha_js+\cfrac{1}{\alpha_{j+1}s+\cfrac{1}{\ddots\,\alpha_{n-1}s+\cfrac{1}{\alpha_ns}}}},$$

$$j=2,\cdots,n,$$

对 $F_1(s)$，上式中的 α_js 改为 $1+\alpha_1s$，常数 α_i,β_i 递推计算. 记 $a_{2k}^0=a_{2k}$，$a_{2k}^1=a_{2k+1}$，$k=0,1,\cdots$，常数 α_i 递推计算如下：

$$a_0^{j+1}=a_2^{j-1}-\alpha_ja_2^j,$$

$$a_2^{j+1}=a_4^{j-1}-\alpha_ja_4^j,$$

$$\cdots\cdots\cdots\cdots$$

$$a_{n-j-2}^{j+1}=a_{n-j}^{j-1}-\alpha_ja_{n-j}^j\quad(j=1,2,\cdots,n-1).$$

若 $n-j$ 是奇数，最后一个方程改为

$$a_{n-j-1}^{j+1}=a_{n-j+1}^{j-1},$$

式中 $\alpha_j=a_0^{j-1}/a_0^j$ $(j=1,2,\cdots,n-1)$. 记

$$b_{2k}^1=b_{2k},\ b_{2k}^2=b_{2(k+1)}\quad(k=0,1,\cdots),$$

$$b_{j-2}^{i+2}=b_j^i-\beta_ia_j^i,$$

$$j=\begin{cases}2,4,\cdots,n-i,n-i\ \text{为偶数},\\1,3,\cdots,n-i-1,n-i\ \text{为奇数}\\(i=1,2,\cdots,n-2),\end{cases}$$

式中 $\beta_i=b_0^i/a_0^i$ $(i=1,2,\cdots,n)$. 记

$$\hat{D}_k(s)=\sum_{i=1}^{k}\beta_i\prod_{j=1}^{i}g_{jk}(s),$$

$$g_{jk}(s)=\cfrac{1}{\alpha_js+\cfrac{1}{\alpha_{j+1}s+\cfrac{1}{\ddots\,\alpha_{k-1}s+\cfrac{1}{\alpha_ks}}}},$$

对 $g_{1k}(s)$，上式中的 α_js 改为 $1+\alpha_1s$；则 $G(s)$的 k 阶劳斯近似为

$$G(s)=\frac{1}{s}\hat{G}_k\left(\frac{1}{s}\right).$$

由于劳斯近似的分母系数仅由原模型 $G(s)$的分母系数确定，所以稳定系统的各阶劳斯近似保持稳定.

连分式近似(continued fraction approximation)　大规模系统的一种频域降阶法. 传递函数 $G(s)$的连分式展开式有三种基本形式：

1. $$G(s)=\cfrac{1}{h_1s+\cfrac{1}{h_2+\cfrac{1}{h_3s+\cfrac{1}{h_4+\cfrac{1}{\ddots}}}}}.$$

2. $$G(s)=\cfrac{1}{k_1+\cfrac{1}{\cfrac{k_2}{s}+\cfrac{1}{k_3+\cfrac{1}{\cfrac{k_4}{s}+\cfrac{1}{\ddots}}}}}.$$

3. $$G(s)=\cfrac{1}{d_1+f_1s+\cfrac{1}{\cfrac{d_2}{s}+f_2+\cfrac{1}{d_3+f_3s+\cfrac{1}{\cfrac{d_4}{s}+f_4+\cfrac{1}{\ddots}}}}}.$$

连分式近似是在 $G(s)$的连分式展开式中舍去某次系数后所获得的低阶模型.

平衡截尾(balanced truncation)　大规模系统的一种降阶方法. 它是从系统的平衡实现中截去那些最小的汉克尔奇异值所对应状态的一种降阶方法. 设(A,B,C)为系统 G 的一个有序平衡实现，指

的是 A 渐近稳定且系统的可控可观格拉姆矩阵相等且都为对角阵，即有

$$A\Sigma+\Sigma A^{\mathrm{T}}+BB^{\mathrm{T}}=0,$$
$$A^{\mathrm{T}}\Sigma+\Sigma A+C^{\mathrm{T}}C=0,$$

其中 $\Sigma=\mathrm{diag}(\sigma_1 I_{r_1},\sigma_2 I_{r_2},\cdots,\sigma_m I_{r_m})$，$\sigma_1>\sigma_2>\cdots>\sigma_m>0$，$n=r_1+r_2+\cdots+r_m$ 为 $C(sI-A)^{-1}B$ 的麦克米兰度，r_i 为 σ_i 的重数，σ_i 为系统的汉克尔奇异值. 令 $\Sigma=\mathrm{diag}(\Sigma_1,\Sigma_2)$，这里

$$\Sigma_1=\mathrm{diag}(\sigma_1 I_{r_1},\cdots,\sigma_l I_{r_1}),$$
$$\Sigma_2=\mathrm{diag}(\sigma_{l+1} I_{r_{l+1}},\cdots,\sigma_m I_{r_m})$$

对应于 Σ 的形式，将 A,B,C 分块如下：

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},\ B=\begin{bmatrix}B_1\\ B_2\end{bmatrix},\ C=[C_1,C_2],$$

这时具有实现 (A_{11},B_1,C_1,D) 的降阶系统 $\hat{G}$ 为 G 的一个平衡截尾. 平衡截尾具有如下性质：

1. 平衡截尾保持稳定性和最小性，即 A_{11} 渐近稳定且 (A_{11},B_1,C_1) 是最小实现.

2. 降阶系统 $\hat{G}$ 具有麦克米兰度 $r=r_1+r_2+\cdots+r_l$，并且逼近误差满足

$$\|G-\hat{G}\|_\infty\leqslant 2(\sigma_{l+1}+\cdots+\sigma_m).$$

H_∞ 模型降阶（H_∞ model reduction） 以 H_∞ 范数为指标的系统模型降阶方法. H_∞ 模型降阶包括最优和次优 H_∞ 模型降阶两种.

最优 H_∞ 模型降阶：对于稳定的 n 阶模型 G，寻找一个稳定的 $r(r<n)$ 阶模型 $\hat{G}$，使得 H_∞ 误差 $\|G-\hat{G}\|_\infty$ 最小.

次优 H_∞ 模型降阶：对于给定的正数 γ，寻找一个稳定的 $r(r<n)$ 阶模型 $\hat{G}$（如果存在），使得

$$\|G-\hat{G}\|_\infty<\gamma.$$

次优 H_∞ 模型降阶，当给定的正数 γ 属于一定范围时，可采用模型降阶中的平衡截尾方法和最优汉克尔范数逼近方法得到一定阶数的降阶模型.

利用线性矩阵不等式技术可以从理论上给出存在 $r(r<n)$ 阶系统 $\hat{G}$ 使得次优 H_∞ 模型降阶问题有解的充分必要条件，并把问题转化为具有非凸的矩阵秩约束条件下的线性矩阵不等式可解问题. 求解这类问题，数值上可采用交替投影的方法，但算法的全局收敛不能保证.

大规模系统的分解（decomposition for large scale system） 大系统控制特有的一种设计方法. 大系统由多个子系统构成，通常维数很高，且子系统间存在着复杂的关联关系. 传统控制理论的方法用于高维大系统时，涉及问题的复杂性使之难以胜任或代价太大，分解与协调方法正是克服这种弊病而产生的一种方法. 分解是指把复杂的整体问题分解为简单的子问题各自求解，当然各子问题分别求得的解一般不是整体问题的解. 协调是指按大系统的整体目标和关联约束来协调各子问题所得到的结果，以求得整体问题的解.

大规模系统的协调（coordination for large scale system） 见“大规模系统的分解”.

模型协调法（model coordination method） 处理大系统问题的一种基本分解、协调方法. 其特点是把子系统间的关联变量作为协调变量，并进行预估. 比起集中计算，该方法的计算量要大为减少. 下面以静态最优化问题为例来介绍模型协调方法的基本过程：

$$\min J(x,u,y) \tag{1}$$

满足条件

$$f(x,u,y)=0, \tag{2}$$

这里 x 是状态向量，u 是控制向量，y 是子系统间的关联向量. 把所讨论的问题和目标函数分解成两个子系统：

$$J(x,u,y)=J_1(x_1,u_1,y_1)+J_2(x_2,u_2,y_2), \tag{3}$$
$$f_i(x_i,u_i,y_1,y_2)=0,\ i=1,2, \tag{4}$$

式中 x_i,u_i,y_i 分别是第 i 个子系统的状态向量、控制向量和关联向量. 这种分解使每个子系统有一个目标函数，然而通过向量 $y_i(i=1,2)$ 使各子系统仍然是关联的. 模型协调法是对关联变量做出预估，即取 $y_i=w_i(i=1,2)$，在这样的条件下问题(1)，(2)可划分成下面两级问题：

第一级问题：对于第 i 个子系统，求

$$K_i(w)=\min_{x_i,u_i} J_i(x_i,u_i,w_i)\ (i=1,2), \tag{5}$$
$$f_i(x_i,u_i,w_1,w_2)=0\ (i=1,2); \tag{6}$$

第二级问题：

$$\min_w K(w)=K_1(w)+K_2(w). \tag{7}$$

分别对下列可行集求上述极小化：

$$S_1^i=\{(x_i,u_i):f_i(x_i,u_i,w)=0\}\ (i=1,2), \tag{8}$$
$$S_2^i=\{w_i:K_i(w_i)\text{存在}\}\qquad (i=1,2). \tag{9}$$

解出的 w_i 就作为 y_i 的新预估值，返回第一级，直到 $|K(w^n)-K(w^{n-1})|<\varepsilon$ 为止. 其中 w^n 为第 n 次迭代的预估值，ε 为所需要的精度. 最后得到大系统的近似最优控制.

目标协调法（goal coordination method） 处理大系统问题的一种基本分解、协调方法. 其特点是通过切断子系统之间的一切联系来严格消除相互作用. 对于静态大系统的最优化问题和动态大系统的控制问题，比起集中计算，该方法也能减少计算量. 以静态最优化问题为例，目标协调方法的基本过程如下.

$$\min J(x,u,y) \tag{1}$$

满足条件

$$f(x,u,y)=0, \tag{2}$$

这里 x 是状态向量，u 是控制向量，y 是子系统间的

关联向量. 引入变量 $z=(z_1,z_2)$，当 $y_i=z_i$ 时，(2)有以下的等价形式：

$$f_1(x_1,u_1,y_1,z_2)=0, \tag{3}$$

$$f_2(x_2,u_2,y_2,z_1)=0. \tag{4}$$

在目标函数中加入一个补偿项，即

$$J(x,u,y,z,\alpha)=J_1(x_1,u_1,y_1,z_2)+J_2(x_2,u_2,y_2,z_1)+\alpha^{\mathrm{T}}(y-z), \tag{5}$$

则可产生下面两级问题：

第一级问题：系统(1)，(2)解耦为两个子系统. 子系统 1 为

$$\min_{(x_1,u_1,y_1,z_2)} J_1(x_1,u_1,y_1,z_2)+\alpha_1^{\mathrm{T}}y_1-\alpha_2^{\mathrm{T}}z_2, \tag{6}$$

满足条件(3)；子系统 2 为

$$\min_{(x_2,u_2,y_2,z_1)} J_2(x_2,u_2,y_2,z_1)-\alpha_1^{\mathrm{T}}z_1+\alpha_2^{\mathrm{T}}y_2, \tag{7}$$

满足条件(4). 将 y_i 看做第 i 个子系统的输出向量，z_i 为输入向量. 两个子系统的联系完全断开时，一般 $y_i\neq z_i$.

第二级问题：控制协调变量 α，使两个子系统的关联误差为零，即

$$\min_{\alpha} e=\min_{\alpha}|y-z|=0,$$

这样使子系统保持关联平衡，即 $y_i=z_i$.

关联预测法(interaction prediction approach) 用于研究线性动态大系统二次型性能指标最优控制问题的一种方法. 主要思想是将子系统间的关联变量及拉格朗日乘子作为协调变量，并分别对其进行预估，所以实际上是一种混合方法. 基本方法是将关联大系统分解成 N 个子系统：

$$\begin{cases}\dot{x}_i(t)=A_ix_i(t)+B_iu_i(t)+C_iz_i(t),\\ x_i(0)=x_{i0}\quad(i=1,2,\cdots,N),\end{cases} \tag{1}$$

式中 $x_i(t)\in\mathrm{R}^{n_i}$，$u_i(t)\in\mathrm{R}^{m_i}$，$z_i(t)\in\mathrm{R}^{n_i}$，$A_i,C_i\in\mathrm{R}^{n_i\times n_i}$，$B_i\in\mathrm{R}^{n_i\times m_i}$，关联向量 z_i 为

$$z_i(t)=\sum_{j=1}^{N}G_{ij}x_j(t). \tag{2}$$

性能指标是加性可分的，即

$$J=\sum_{i=1}^{N}J_i=\sum_{i=1}^{N}\left(\frac{1}{2}x_i^{\mathrm{T}}(t_{\mathrm{f}})P_ix_i(t_{\mathrm{f}})+\frac{1}{2}\int_0^{t_{\mathrm{f}}}\{x_i^{\mathrm{T}}(t)Q_ix_i(t)+u_i^{\mathrm{T}}(t)R_iu_i(t)\}\mathrm{d}t\right). \tag{3}$$

第一级问题：第 i 个子问题为

$$\min J_i \tag{4}$$

并满足约束条件(1)，(2). 引入一组拉格朗日乘子 $\alpha_i(t)$ 和同态向量 $p_i(t)$，得第 i 个子系统的哈密顿算子

$$H_i=\frac{1}{2}x_i^{\mathrm{T}}(t)Qx_i(t)+\frac{1}{2}u_i^{\mathrm{T}}(t)R_iu_i(t)+\alpha_i^{\mathrm{T}}x_i-\sum_{j=1}^{N}\alpha_j^{\mathrm{T}}G_{ji}x_i+p_i^{\mathrm{T}}(A_ix_i+B_iu_i+C_ix_i). \tag{5}$$

求解(5)得 $x_i(t)$，$u_i(t)$，$p_i(t)$.

第二级问题：校正新的协调变量$[\alpha_i^{\mathrm{T}}(t),z_i^{\mathrm{T}}(t)]^{\mathrm{T}}$对第$(l+1)$次迭代：

$$\begin{pmatrix}\alpha_i(t)\\ z_i(t)\end{pmatrix}^{l+1}=\begin{bmatrix}-C_i^{\mathrm{T}}p_i(t)\\ \sum_{j=1}^{N}G_{ij}x_j(t)\end{bmatrix}^{l}. \tag{6}$$

返回第一级，直到

$$|z_i^{l+1}(t)-z_i^{l}(t)|<\delta,\ |\alpha_i^{l+1}(t)-\alpha_i^{l}(t)|<\delta \tag{7}$$

为止. 其中 δ 为要求精度.

关联预测法的突出优点是不必计算复杂的梯度及搜索最优步长，而只要逐次迭代即可.

协态预测法(costate prediction approach) 处理非线性动态大系统二次型性能指标最优控制问题的有效方法之一. 主要思想是把状态和目标函数中的线性可分部分隔离开来，并对其非线性及不可分的部分通过预估加以相对固定. 考虑非线性动态大系统

$$\dot{x}=f(x,u,t), \tag{1}$$

性能指标为如下的二次型：

$$\min J=\frac{1}{2}x^{\mathrm{T}}(t_{\mathrm{f}})Px(t_{\mathrm{f}})+\frac{1}{2}\int_0^{t_{\mathrm{f}}}(x^{\mathrm{T}}(t)Qx(t)+u^{\mathrm{T}}(t)Ru(t))\mathrm{d}t, \tag{2}$$

其中 $x\in\mathrm{R}^n$，$u\in\mathrm{R}^m$，P,Q 半正定，R 正定. 把(1)，(2)中的线性可分部分隔离开来. 若取预估值 $x=x^*$，$u=u^*$，则可得

$$\dot{x}=A(x^*,u^*,t)x+B(x^*,u^*,t)u+C(x^*,u^*,t), \tag{3}$$

$$\min J=\frac{1}{2}x^{\mathrm{T}}(t_{\mathrm{f}})P_1x(t_{\mathrm{f}})+\frac{1}{2}x^{*\mathrm{T}}(t_{\mathrm{f}})P_2x^*(t_{\mathrm{f}})+\frac{1}{2}\int_0^{t_{\mathrm{f}}}(x^{\mathrm{T}}(t)Qx(t)+x^{*\mathrm{T}}Qx^*(t)+u^{\mathrm{T}}(t)R_1u(t)+u^{*\mathrm{T}}(t)R_2u^*(t))\mathrm{d}t, \tag{4}$$

其中 A,B,P_1,Q_1,R_1 为分块对角阵，其第 N 块对应于第 N 个子系统，且 P_{1i},Q_{1i}常正，R_{1i}正定$(i=1,2,\cdots,N)$，$P_1+P_2=P$，$Q_1+Q_2=Q$，$R_1+R_2=R$.

$$A(x,u,t)x+B(x,u,t)u+C(x,u,t)=f(x,u,t). \tag{5}$$

这一问题的哈密顿函数为

$$H=\frac{1}{2}x^{\mathrm{T}}(t)Q_1x(t)+\frac{1}{2}x^{*\mathrm{T}}(t)Qx^*(t)+\frac{1}{2}u^{\mathrm{T}}(t)R_1x(t)+\frac{1}{2}u^{*\mathrm{T}}(t)R_2u^*(t)+\lambda^{\mathrm{T}}(Ax+Bu+C)+\beta^{\mathrm{T}}(x-x^{\mathrm{T}})+\gamma(u-u^*). \tag{6}$$

第一级问题：给定了预估的 x^*，u^*，γ，λ 后，由于 A,B,P_1,Q_1,R_1 均为分块对角阵，所以可分解为各子系统独立求解，得 x,u.

第二级问题：将第一级算出的结果作为协调向量

$$e = [x^{*T}, u^{*T}, \lambda^T, \gamma^T]^T \quad (7)$$

的新预估值. 迭代结束的标志为

$$\int_0^t \| e^{l+1} - e^l \| dt < \varepsilon,$$

其中 l 为迭代次数, ε 为给定的精度.

多级递阶控制(multilevel hierarchical control) 有分层结构的大系统的一种控制方式. 设系统分为多层, 第 i 层有如下优化问题:

目标函数: $J(x,u,y)$;

约束方程: $f(x,u,y)=0$;

x,u 分别是第 i 层的状态变量和控制变量, y 是第 $i+1$ 层子系统间的耦合变量. 第 $i+1$ 层各子系统有优化问题:

目标函数: $J_k(x^k,u^k,y^k)$;

约束方程: $f_k(x^k,u^k,y)=0 \ (k=1,2,\cdots,m)$;

其中 x^k,u^k,y^k 分别是第 $i+1$ 层第 k 个子系统的状态、控制和耦合变量. 上下两层目标函数满足

$$J(x,u,y) = \sum_{k=1}^{m} J_k(x^k,u^k,y^k).$$

多级递阶控制将第 i 层优化问题化为下面两个子问题:

1. 固定或不考虑耦合变量, 使第 $i+1$ 层各子系统独立, 解各子系统优化问题.

2. 考虑耦合项, 在第 i 层协调各子系统的目标函数值, 求得第 i 层目标函数最优解.

结构摄动法(structural perturbation approach) 线性动态大系统递阶控制的方法之一. 设大系统可用 N 个子系统描述:

$$\begin{cases} \dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + \sum_{j=1}^{N} G_{ij} x_j(t), \\ x_i(t_0) = x_{i_0} \quad (i=1,2,\cdots,N), \end{cases} \quad (1)$$

其中

$$x_i(t) \in \mathrm{R}^{n_i},\ u_i(t) \in \mathrm{R}^{m_i},\ \sum_{i=1}^{N} n_i = n,\ \sum_{i=1}^{N} m_i = m,$$

A_i, B_i, G_{ij} 为适当维数的常数矩阵. 先忽略关联项, 求出各解耦子系统

$$\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t), i = 1,2,\cdots,N \quad (2)$$

在子系统性能指标

$$J_i = \frac{1}{2}\int_{t_0}^{\infty} [x_i^T(t) Q_i x_i(t) + u_i^T(t) R_i u_i(t)] dt$$

下的最优控制

$$u_i^*(t) = -R_i^{-1} B_i^T K_i x_i(t) \triangleq -P_i x_i(t), \quad (3)$$

其中 K_i 是第 i 个子系统的黎卡提矩阵代数方程

$$A_i^T K_i + K_i A - K_i B_i R_i^{-1} B_i^T K_i + Q_i = 0$$

的对称正定解. 记在控制(3)下各解耦子系统(2)的性能指标 J_i 的最优值为 J_i^*, 这时各解耦子系统最优性能指标之和

$$J^* = \sum_{i=1}^{N} J_i^*$$

与大系统整体的最优性能指标 J 之间有三种关系:

当 $J=J^*$ 时, 称子系统间的关联是中性的;

当 $J<J^*$ 时, 称关联是有利的;

当 $J>J^*$ 时, 称关联是无利的.

记 $G \triangleq (G_{ij})$, $K \triangleq \mathrm{blockdiag}(K_1,K_2,\cdots,K_N)$, 可以证明, 当且仅当存在反对称阵 S, 使得 $G=SK$ 时, 关联才是中性的. 显然, 对于一般大系统, 关联矩阵 G 并不一定满足这一特殊条件. 这时可以再采用来自上一级的全局控制

$$u_i^g = -\sum_{j=1}^{N} H_{ij} x_j(t) \quad (i=1,\cdots,N)$$

来改进大系统整体的性能指标. 若记

$A = \mathrm{blockdiag}\ (A_1,A_2,\cdots,A_N)$,

$B = \mathrm{blockdiag}\ (B_1,B_2,\cdots,B_N)$,

$P = \mathrm{blockdiag}\ (P_1,P_2,\cdots,P_N)$,

$H = \{H_{ij}\}$, $x(t) = (x_1^T,x_2^T,\cdots,x_N^T)^T$,

$u(t) = (u_1^T,u_2^T,\cdots,u_N^T)^T$,

则总的控制律可写为 $u(t) = -(P+H)x(t)$. 若全局修正增益阵 H 满足 $BH=B^pP$, 则闭环大系统可表示为

$$\dot{x}(t) = (A+G)x(t) - (B+B^p)Px(t). \quad (4)$$

所谓结构摄动法, 是要从关联项 G 确定控制矩阵 B 的结构摄动值 B^p, 并估计性能指标变化的界. 其主要研究结果为: 若 $A+G$ 非奇异, 反对称矩阵 S 是李亚普诺夫矩阵方程

$$S(A+G)+(A+G)^T S + G^T KA - A^T KG = 0$$

的解, 且使

$$K + (S-KG)(A+G)^{-1} \triangleq K + \hat{K}$$

为正定矩阵, 则输入矩阵的摄动值

$$B^p = -(K+\hat{K})^{-1}\hat{K}B,$$

闭环大系统(4)的性能指标

$$\hat{J} = \frac{1}{2} x^T(t_0)(K+\hat{K})x(t_0),$$

且性能指标偏差

$$\rho = \frac{\hat{J} - J^*}{J^*} \leqslant \frac{\lambda_M(\hat{K})}{\lambda_m(K)},$$

这里 $\lambda_m(\cdot)$ 及 $\lambda_M(\cdot)$ 分别表示有关矩阵特征值的极小值和极大值.

奇异摄动的多重时标分解(multiple time-scale decomposition for singular perturbation) 按不同时间尺度对系统进行分解的方法. 对于呈现快运动和慢运动的不同时标的大系统, 可按其时标进行分解, 以简化设计和计算. 一个双时标的大系统 Σ, 利用奇异摄动方法可分离出快时标系统 Σ_f 和慢时标系统 Σ_s, 原系统的目标函数 J 也可相应地分解为 Σ_f 的目标函数 J_f 和 Σ_s 的目标函数 J_s. 对 (Σ_f, J_f) 和

(Σ_s, J_s)这两个独立的低维最优控制问题，可分别求得优化的控制阵，然后通过还原算法，最终可得到整个系统的次优控制. 这一按时标分离的控制算法可推广到多时标系统. 这种奇异摄动的多重时标分解与其他按空间或时间方向分解的递阶算法之间有着如下重要差别：

1. 该分解深入到系统动态的本质方面，分解后所得到的快、慢子系统是分离的，相应的子问题也是相互独立的.

2. 快、慢子系统模型只是原系统模型的近似，因此导出的控制律也只能是原问题的次优解.

大系统的稳定性分解(stability decomposition for large scale system)　简化大系统稳定性分析的方法. 由于大系统一般阶数较高，不便于直接对它进行稳定性分析. 大系统的稳定性分解是指在分析大系统稳定性时，将其分解为若干子系统，通过对各孤立子系统的稳定性分析和对子系统间关联项性质的研究，得出判别大系统稳定性的充分条件(大系统的稳定性分解有多种方法，如按各子系统的李亚普诺夫函数进行分解、按大系统的结构进行分解等). 利用大系统的稳定性分解可简化大系统的稳定性分析，但所得结果一般较为保守.

李亚普诺夫函数分解法(decomposition method by Liapunov function)　大系统稳定性分解的一种方法. 其一般步骤为：

1. 将大系统分解为若干子系统，使得可找出各孤立子系统的李亚普诺夫函数，并用它分析各孤立子系统的稳定性.

2. 研究子系统间关联项的性质，通常是估计出各关联项的上界.

3. 利用各孤立子系统的李亚普诺夫函数和关联项的性质，得出判定大系统稳定性的条件.

在大系统稳定性分析中最常用的李亚普诺夫函数分解法是(加权和)标量李亚普诺夫方法和向量李亚普诺夫方法.

加权和标量李亚普诺夫方法(method of weighted sum for scalar Liapunov function)　一种常用的用李亚普诺夫函数分解法分析大系统稳定性的方法，常简称加权和李亚普诺夫方法或标量李亚普诺夫方法. 其典型作法可简述如下：研究大系统

$$\dot{x} = f(x,t), \qquad (1)$$

式中$x\in \mathrm{R}^n$，f连续，保证大系统(1)在$J\triangleq[t_0,\infty)$上存在惟一解，且$x=0$是系统的惟一平衡点. 设它可分解为m个$(m<n)$子系统

$$\dot{y}_i = f_i(y_i,t) + g_i(y_1,y_2,\cdots,y_m,t) \qquad (2)$$
$$(i=1,2,\cdots,m),$$

式中$x=[y_1^{\mathrm{T}},y_2^{\mathrm{T}},\cdots,y_m^{\mathrm{T}}]^{\mathrm{T}}$，$y_i\in \mathrm{R}^{n_i}$，$f_i$和$g_i$保证子系统(2)在$J$上存在惟一解，又当$y_i=0$时，$f_i(y_i,t)\equiv 0$. 如果：

1. 对每个孤立子系统

$$\dot{y}_i = f_i(y_i,t) \quad (i=1,2,\cdots,m) \qquad (3)$$

都存在连续可微函数$v_i(y_i,t)$，满足

$$\begin{cases}\varphi_{i1}(\|y_i\|)\leqslant v_i(y_i,t)\leqslant\varphi_{i2}(\|y_i\|),\\ \dot{v}_i(y_i,t)|_{(3)}\leqslant\sigma_i\varphi_{i3}(\|y_i\|)\end{cases}$$
$$(\forall y_i\in\mathrm{R}^{n_i},\ \forall t\in J),$$

其中φ_{i1}，φ_{i2}为KR类函数，φ_{i3}为K类函数，σ_i为常数；

2. 每个子系统(2)的关联项$g_i(y_1,y_2,\cdots,y_m,t)$满足：存在常数$a_{ij}(j=1,2,\cdots,m)$，使得

$$[\nabla v_i(y_i,t)]^{\mathrm{T}}g_i(y_1,y_2,\cdots,y_m,t)$$
$$\leqslant\sqrt{\varphi_{i3}(\|y_i\|)}\sum_{j=1}^{m}a_{ij}\sqrt{\varphi_{j3}(\|y_j\|)},$$
$$\forall\,[y_1^{\mathrm{T}},y_2^{\mathrm{T}},\cdots,y_m^{\mathrm{T}}]^{\mathrm{T}}\in\mathrm{R}^n,\ \forall t\in J;$$

则可取各孤立子系统$v_i(y_i,t)$的加权和作为大系统(1)的李亚普诺夫函数v，即取

$$v(x,t) = \sum_{i=1}^{m}\alpha_i v_i(y_i,t),$$

式中$\alpha_i(i=1,2,\cdots,m)$为正的待定加权系数. 利用条件1和2可算出v沿大系统(1)解的导数

$$\dot{v}(x,t)|_{(1)}\leqslant -w^{\mathrm{T}}\left[\frac{1}{2}(\Lambda D+D^{\mathrm{T}}\Lambda)\right]w,$$

式中

$$w=\left[\sqrt{\varphi_{13}(\|y_1\|)},\cdots,\sqrt{\varphi_{m3}(\|y_m\|)}\right]^{\mathrm{T}},$$
$$\Lambda=\mathrm{diag}[\alpha_1,\alpha_2,\cdots,\alpha_m],$$
$$D=[d_{ij}],\ d_{ij}=\begin{cases}-(\sigma_i+a_{ii}) & (i=j),\\ -a_{ij} & (i\neq j).\end{cases}$$

由李亚普诺夫稳定性理论知，若

$$\frac{1}{2}(\Lambda D+D^{\mathrm{T}}\Lambda)$$

正定，则大系统(1)的零解是全局一致渐近稳定的. 用这种方法可以把n阶非线性大系统的稳定性分析转化为判定m阶$(m<n)$常值矩阵正定性的问题，使问题大大简化. 又若条件2中的a_{ij}均非负，则可利用正定阵与M矩阵(参见“M矩阵”)的关系得到：若D为M矩阵，则大系统的零解是全局渐近稳定的. 这一稳定性判据与待定的加权系数无关，使用更加方便.

向量李亚普诺夫函数(vectorial Liapunov function)　分析大系统稳定性的一个重要向量值函数. 它沿着所考虑的大系统的解对时间的全导数满足一个向量微分不等式. 在一定条件下，这一向量微分不等式对应的低维辅助系统的稳定性蕴涵原高维系统的相应稳定性. 因此向量李亚普诺夫函数的作用是把一个高维系统的稳定性问题简化为一个低维集结系统的稳定性问题.

比较原理(comparison principle) 反映满足(向量)微分不等式的向量值函数与对应于微分不等式的微分方程的最大(或最小)解之间的一种不等式关系. 它通常表示为如下形式的定理.

定理:设 l 维向量 $v(t)$ 满足向量微分不等式

$$\dot{v} \leqslant w(t,v) \text{ (或 } \dot{v} \geqslant w(t,v)\text{)}, \tag{1}$$

其中向量值函数 $w=(w_1,w_2,\cdots,w_l)^{\mathrm{T}}:\mathrm{R}_+ \times \mathrm{R}^l \to \mathrm{R}^l$ 是拟单调增的,即对于每个 $w_j, j=1,2,\cdots,l$,只要 $\alpha_i \leqslant \beta_i (i \neq j)$ 及 $\alpha_j=\beta_j$,就有 $w_j(t,\alpha) \leqslant w_j(t,\beta)$,这里 $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_l)$, $\beta=(\beta_1,\beta_2,\cdots,\beta_l)$. 记 $r_{\mathrm{M}}(t)$ (或 $r_{\mathrm{m}}(t)$)为方程

$$\dot{r} = w(t,r) \tag{2}$$

满足 $r(t_0)=r_0$ 并定义在 $t \geqslant t_0$ 上的最大解(或最小解),则当 $v(t_0) \leqslant r(t_0)$ (或 $v(t_0) \geqslant r(t_0)$)时,对所有 $t \geqslant t_0$ 成立 $v(t) \leqslant r(t)$ (或 $v(t) \geqslant r(t)$). 上面的所有向量不等式应理解为对应分量之间的不等式.

向量李亚普诺夫方法(method of vectorial Liapunov function) 一种常用的大系统稳定性分析方法. 其理论基础是(向量形式的)微分方程比较原理. 它的典型作法可简述如下:研究大系统

$$\dot{x} = f(x,t), \tag{1}$$

式中 $x \in \mathrm{R}^n$, f 连续,保证大系统(1)在 $J \triangleq [t_0,\infty)$ 上存在惟一解,且 $x=0$ 是系统的惟一平衡点. 设它可分解为 m 个($m<n$)子系统

$$\dot{y}_i = f_i(y_i,t) + g_i(y_1,y_2,\cdots,y_m,t) \tag{2}$$
$$(i = 1,2,\cdots,m),$$

式中 $x=[y_1^{\mathrm{T}},y_2^{\mathrm{T}},\cdots,y_m^{\mathrm{T}}]^{\mathrm{T}}$, $y_i \in \mathrm{R}^{n_i}$, f_i 和 g_i 保证子系统(2)在 J 上存在惟一解,又当 $y_i=0$ 时, $f_i(y_i,t) \equiv 0$. 如果:

1. 对每个孤立子系统

$$\dot{y}_i = f_i(y_i,t) \quad (i = 1,2,\cdots,m), \tag{3}$$

都存在连续可微函数 $v_i(y_i,t)$ 满足

$$\begin{cases} \varphi_{i1}(\| y_i \|) \leqslant v_i(y_i,t) \leqslant \varphi_{i2}(\| y_i \|), \\ \dot{v}_i(y_i,t)|_{(3)} \leqslant \sigma_i v_i(y_i,t) \\ (\forall y_i \in \mathrm{R}^{n_i},\ \forall t \in J), \end{cases} \tag{4}$$

其中 $\varphi_{i1},\varphi_{i2}$ 为 KR 类函数, σ_i 为常数;

2. 每个子系统(2)的关联项 $g_i(y_1,y_2,\cdots,y_m,t)$ 满足:存在常数 $a_{ij}(j=1,2,\cdots,m)$,使得

$$[\nabla v_i(y_i,t)]^{\mathrm{T}} g_i(y_1,y_2,\cdots,y_m,t)$$
$$\leqslant \sqrt{v_i(y_i,t)} \sum_{j=1}^{m} a_{ij} \sqrt{v_j(y_j,t)}, \tag{5}$$
$$\forall [y_1^{\mathrm{T}},y_2^{\mathrm{T}},\cdots,y_m^{\mathrm{T}}]^{\mathrm{T}} \in \mathrm{R}^n,\ \forall t \in J,$$

定义

$$D = [d_{ij}],\ d_{ij} = \begin{cases} -(\sigma_i + a_{ii}) & (i = j), \\ -a_{ij} & (i \neq j); \end{cases}$$

则当 D 为 M 矩阵时,可由(4)中第二式和(5)式算出

$$\frac{\mathrm{d}}{\mathrm{d}t} \sqrt{v_i(y_i,t)} \Big|_{(2)} \leqslant \frac{1}{2} \sum_{j=1}^{m} d_{ij} \sqrt{v_j(y_j,t)} \tag{6}$$
$$(i = 1,2,\cdots,m).$$

引入向量李亚普诺夫函数

$$V = \left[\sqrt{v_1(y_1,t)},\cdots,\sqrt{v_m(y_m,t)}\right]^{\mathrm{T}}$$

后,可将(6)式写为

$$\dot{V}|_{(1)} \leqslant -\frac{1}{2}DV.$$

因由(4)的第一式知存在 KR 类函数 φ_1,φ_2,使得

$$\varphi_1(\| x \|) \leqslant \| V(x,t) \| \leqslant \varphi_2(\| x \|),$$

又由 D 为 M 矩阵知, $-(1/2)DV$ 是 V 的拟单调增函数,且 $-(1/2)D$ 是稳定阵,故由(6)式和微分方程比较原理可证明大系统(1)的零解是全局一致渐近稳定的.

上述向量李亚普诺夫函数方法把 n 阶非线性大系统的稳定性分析转化为判断一个 m 阶($m<n$)常数阵是否为 M 矩阵的问题,简化了大系统的稳定性分析.

M 矩阵(M-matrix) 两种特殊的方阵. 在大系统稳定性分析中用到两种不同的 M 矩阵定义. 一种是闵科夫斯基矩阵. 若实方阵 $A=[a_{ij}]$ 的非对角元均非正(即 $a_{ij} \leqslant 0, i \neq j$),且其所有主子式均为正,则称其为闵科夫斯基矩阵,简称 M 矩阵. M 矩阵有许多等价的定义,常用的有:对每一个非对角元均非正的实方阵 A, A 是 M 矩阵 $\Leftrightarrow A$ 的所有主要主子式均为正 $\Leftrightarrow A$ 的所有特征值实部均为正 $\Leftrightarrow$ 存在向量 $\alpha>0$,使得 $A\alpha>0 \Leftrightarrow$ 存在向量 $\alpha>0$,使得 $\alpha^{\mathrm{T}}A>0 \Leftrightarrow A$ 可逆且 A^{-1} 的所有元素均非负. M 矩阵与正定阵间有重要关系. 若实方阵 A 的非对角元均非正,则 A 是 M 矩阵的充分必要条件是:存在正对角阵 Λ,使得

$$\frac{1}{2}(\Lambda A + A^{\mathrm{T}}\Lambda)$$

是正定阵. 在用加权和标量李亚普诺夫函数方法分析大系统稳定性时,可用这一关系使所得到的判据与李亚普诺夫函数中的加权系数无关.

另一种是麦茨勒尔矩阵. 若方阵 $A=[a_{ij}]$ 所有非对角元均非负(即 $a_{ij} \geqslant 0, i \neq j$),则称之为麦茨勒尔矩阵,也简称 M 矩阵. 一个麦茨勒尔矩阵稳定(即其特征值实部均为负)的充分必要条件是:它的所有奇数阶主要主子式均为负,所有偶数阶主要主子式均为正.

上述两种 M 矩阵间有如下关系: A 是闵科夫斯基矩阵等价于 A 是稳定的麦茨勒尔矩阵. 在遇到 M 矩阵时要注意区别这两种不同的定义. 利用加权和标量李亚普诺夫函数方法或向量李亚普诺夫函数方法分析大系统的稳定性时,通常都归结为判定一个矩阵是否是闵科夫斯基矩阵或是否是稳定的麦茨勒

尔矩阵.

闵科夫斯基矩阵(Minkowski matrix) 见“M矩阵”.

麦茨勒尔矩阵(Metzler matrix) 见“M 矩阵”.

连结稳定性(connected stability) 又称为关联稳定性.大系统结构变化时的一种稳定性描述.大系统可以看成由若干个子系统相互耦合而成.当系统结构变化时,耦合关系的强度会发生变化,连结稳定性就是研究系统在这种结构扰动下的稳定性质,其数学描述如下:

$$\dot{x}_i = f_i(t,x_i) + g_i(t,e_{i1}x_1,\cdots,e_{iN}x_N), \quad (i=1,2,\cdots,N), \tag{1}$$

其中

$$f_i:\mathrm{R}_+\times \mathrm{R}^{n_i} \to \mathrm{R}^{n_i},$$
$$g_i:\mathrm{R}_+\times \mathrm{R}^{n_1}\times\cdots\times \mathrm{R}^{n_N},$$
$$e_{ij}:\mathrm{R}_+\to[0,1]$$

是连续函数,$E=[e_{ij}]$是由基本互联矩阵 $\bar{E}=[\bar{e}_{ij}]$生成的互联矩阵,即有 $e_{ij}(t)\leqslant\bar{e}_{ij}(i,j=1,2,\cdots,N)$对 $t\in\mathrm{R}_+$成立,记为 $E\in\bar{E}$,其中 $\bar{e}_{ij}$定义为

$$\bar{e}_{ij}=\begin{cases}1, x_j\text{ 出现在 } g_i(t,\bar{e}_{i1}x_1,\cdots,\bar{e}_{ij}x_j,\\ \quad\cdots,\bar{e}_{iN}x_N)\text{中},\\ 0,\ x_j\text{ 不出现在 } g_i(t,\bar{e}_{i1}x_1,\cdots,\bar{e}_{ij}x_j,\\ \quad\cdots,\bar{e}_{iN}x_N)\text{中}.\end{cases}$$

记

$$\sum_{i=1}^{N}n_i = n,\ x=[x_1^{\mathrm{T}},x_2^{\mathrm{T}},\cdots,x_N^{\mathrm{T}}]^{\mathrm{T}},$$

并通过引入适当的记号,可把大系统(1)表示为下面的等价形式

$$\dot{x} = F(t,x,E). \tag{2}$$

设函数 F 对所有 $E\in\bar{E}$ 均保证(2)(或(1))存在惟一解 $x(t;t_0,x_0,E)(t\geqslant t_0)$满足 $x(t_0;t_0,x_0,E)=x_0$,并设 $x=0$ 是(2)的惟一平衡态.如果对任给的 $\varepsilon>0$ 和 $t_0\in\mathrm{R}_+$,存在 $\delta=\delta(\varepsilon,t_0)>0$,使得 $\|x_0\|<\delta$ 蕴涵 $\|x(t;t_0,x_0,E)\|<\varepsilon$ 对所有 $t\geqslant t_0$ 和 $E\in\bar{E}$ 成立,则称大系统(2)(或(1))的平衡态 $x=0$ 是关联稳定的.

大规模系统的镇定(stabilization of large scale system) 用动态补偿器使闭环大系统的平衡状态为渐近稳定的控制方法.由于大规模系统通常具有分散控制结构,于是如下的分散化镇定方式成为大规模系统镇定的主要特征,即整个动态补偿器由几个独立的补偿器组成,每个补偿器仅以系统的部分输出作为其输入.对于线性时不变大系统,存在一组局部动态补偿器,使闭环系统稳定的充分必要条件是:在相应的分散信息结构下,系统的固定模位于左半开复平面.如果虚轴上或右半开复平面上存在固定模,那么为了镇定系统,必须改变信息结构或采用时变动态补偿器.

分散控制系统(decentralized control system) 信息结构分散的一类大规模系统.设系统表示为

$$\begin{cases}\dot{x} = Ax + \sum_{i=1}^{n}B_iu_i,\\ y_i = C_ix\ \ (i=1,2,\cdots,n),\end{cases}$$

x 是系统状态,y_i,u_i 分别是第 i 个控制站的输出和输入.分散控制系统中,控制器只允许有下面形式:

$$\begin{cases}\dot{z}_i = H_iz_i + L_iy_i,\\ u_i = M_iz_i + F_iy_i + G_iv_i\ \ (i=1,2,\cdots,n),\end{cases}$$

即第 i 个站的控制量只依赖本站的输出量.上述系统是线性的,这个概念也可推广到非线性系统.

分散系统的能控性(controllability of decentralized system) 分散信息结构下的系统能控性.设分散控制系统为

$$\begin{cases}\dot{x} = Ax + \sum_{i=1}^{n}B_iu_i,\\ y_i = C_ix\ \ (i=1,2,\cdots,n),\end{cases}$$

$$I_i(t) = \{y_i(p),u_i(q),\ p\in[0,t],\ q\in[0,t)\}$$

这里 $I_i(t)$是第 i 个站可采用的数据集.对每个控制站,系统矩阵 A,B_i,C_i 已知,若对任意初始状态向量 $x(0)$,存在正数 T 和分散控制律

$$u_i(t) = F_{it}(I_i(t))\ \ (i=1,2,\cdots,n),$$

使 $x(T)=0$,则称系统在分散信息结构下能控.若系统在分散信息结构下能控,则可用局部动态反馈律任意配置闭环极点,反之不成立.

互补子系统(complementary subsystem) 分散控制系统研究中的基本概念.设分散控制系统 Σ 表示为

$$\begin{cases}\dot{x} = Ax + \sum_{i=1}^{N}B_iu_i,\\ y_i = C_ix\ \ (i\in \underline{N}),\end{cases}$$

$\underline{N}$是由 $1,2,\cdots,N$ 构成的集合,$S_*=\{i_1,i_2,\cdots,i_q\}$是 $\underline{N}$ 的真子集,$\bar{S}_*$是 S_*在 $\underline{N}$ 中的补集,

$$B_{S_*}=(B_{i_1},B_{i_2},\cdots,B_{i_q}),$$
$$C_{S_*}^{\mathrm{T}}=(C_{i_1}^{\mathrm{T}},C_{i_2}^{\mathrm{T}},\cdots,C_{i_q}^{\mathrm{T}}),$$

则称$(C_{\bar{S}_*},A,B_{S_*})$为 Σ 的一个互补子系统.若 $j\in S_*(\bar{S}_*)$,则称其为包含输入(输出)通道 j 的一个互补子系统.

强关联系统(strongly connected system) 分散控制系统研究中的基本概念.它是所有互补子系统的传递函数都不为零的分散控制系统.系统的强关联性是通过分散输出反馈使闭环系统变成单通道能控能观的必要条件.

单通道能观性(single channel observability) 分散控制系统研究中的基本概念.它是利用分散输

出反馈使指定控制站对整个系统可观测. 设分散系统为

$$\begin{cases}\dot{x} = Ax + \sum_{i=1}^{n} B_i u_i, \\ y_i = C_i x \quad (i = 1,2,\cdots,n).\end{cases}$$

若存在反馈律 $u_i = K_i y_i \quad (i=1,2,\cdots,n)$,使

$$\left(C_j, A + \sum_{i=1}^{n} B_i K_i C_i\right)$$

为可观测,则称控制站 j 可由分散反馈观测. 设 $\underline{n}$ 是正整数 $1,2,\cdots,n$ 的集合,$S_* = (i_1, i_2, \cdots, i_q)$,$i_j, j \in q$ 互不相同,S_* 是 $\underline{n}$ 的真子集,$\bar{S}_*$ 是 S_* 在 $\underline{n}$ 中的补集. 记 $B_{S_*} = (B_{i_1}, B_{i_2}, \cdots, B_{i_q})$,$C_{S_*}^{\mathrm{T}} = (C_{i_1}^{\mathrm{T}}, C_{i_2}^{\mathrm{T}}, \cdots, C_{i_q}^{\mathrm{T}})$. 若 $j \in \bar{S}_*$,称 (C_{S_*}, A, B_{S_*}) 是系统包含输出通道 j 的一个互补子系统.

用 (C,A,B) 表示一个系统,$A \in \mathrm{R}^{n\times n}$,$B \in \mathrm{R}^{n\times m}$,$C \in \mathrm{R}^{l\times n}$,

$$P(\lambda) = \begin{pmatrix} \lambda I - A & B \\ C & 0 \end{pmatrix},$$

$P(\lambda)$ 的一切 n 阶子式的最高公因式称为 (C,A,B) 的剩余多项式,记为 $\rho(C,A,B)$. 控制站 j 可由分散反馈观测的充分必要条件是:

1. $\left(\begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix}, A\right)$ 是可观测的.

2. 对包含输出通道 j 的一切互补子系统 (C_{S_*}, A, B_{S_*}) 有 $C_{S_*}(\lambda I - A)^{-1} B_{S_*} \not\equiv 0$ 且 $\rho(C_{S_*}, A, B_{S_*}) = 1$.

单通道能控性可类似定义,两者在分散极点配置问题中有重要作用.

固定多项式(fixed polynomial) 分散控制系统闭环特征多项式中不受反馈影响的部分. 设分散控制系统为 Σ:$\{C_i, A, B_i, i \in \underline{N}\}$,记

$$\Omega = \{F \mid F = \mathrm{diag}\{F_1, F_2, \cdots, F_N\}\}$$

是由分散反馈增益矩阵构成的集合,$B = (B_1, B_2, \cdots, B_N)$,$C^{\mathrm{T}} = (C_1^{\mathrm{T}}, C_2^{\mathrm{T}}, \cdots, C_N^{\mathrm{T}})$,$A_F = A + BFC$,对一切 $F \in \Omega$,称 $\det(\lambda I - A_F)$ 的最高公因式为 Σ 关于 Ω 的固定多项式. 固定多项式由分散反馈方式确定,而与分散反馈增益阵的选取无关.

固定模(fixed mode) 分散控制系统研究中的一个重要概念. 指系统中用常增益分散输出反馈控制所不能改变的系统极点. 设分散系统表示为

$$\begin{cases}\dot{x} = Ax + \sum_{i=1}^{n} B_i u_i, \\ y_i = C_i x \quad (i = 1,2,\cdots,n),\end{cases}$$

在每一个控制站,利用反馈律 $u_i = K_i y_i$,K_i 是常数矩阵. 记

$$\mathscr{K} = \{K \mid K = \text{block diag } (K_1, K_2, \cdots, K_n)\},$$

$$C = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix},\ B = [B_1, B_2, \cdots, B_n],$$

$$A_k = A + \sum_{i=1}^{n} B_i K_i C_i = A + BKC.$$

对一切 $K \in \mathscr{K}$,行列式 $|\lambda I - A_k|$ 的最高公因式记为 $\psi(C,A,B,\mathscr{K})$,它称为系统关于 $\mathscr{K}$ 的固定多项式. 固定多项式的根称为固定模. 固定模依赖于反馈信息结构,而与反馈增益阵的取值无关. 上述定义中,第 i 个控制站的控制 u_i 只利用本站信息 y_i. 实际问题中也可采用其他信息结构,固定模随之变化. 固定模在分散控制系统中有极重要的作用. 系统固定模在分散动态反馈中也不能被改变,它是集中系统不可控模和不可观测模的推广.

分散系统的极点配置(pole placement for decentralized system) 大系统中一种控制设计方法. 指利用分散反馈律配置闭环系统极点. 设系统表示为

$$\begin{cases}\dot{x} = Ax + \sum_{i=1}^{N} B_i u_i, \\ y_i = C_i x \quad (i = 1,2,\cdots,N).\end{cases}$$

控制器由 N 个分散控制器组成

$$\begin{cases}\dot{z} = H_i z_i + L_i F_i, \\ u_i = M_i z_i + F_i y_i + G_i v_i \quad (i = 1,2,\cdots,N).\end{cases}$$

复数共轭成对出现的复数集称为对称复数集. $\{q_1, q_2, \cdots, q_r\}$ 是正整数集,Λ 是对称复数集,Λ 可划分为 r 个互不相交的子集 Λ_i,Λ_i 是含有 q_i 个复数的对称集,则称 Λ 和 $\{q_1, q_2, \cdots, q_r\}$ 相容. 若存在某个正整数集 $\{q_1, q_2, \cdots, q_r\}$,对于和它相容的对称复数集 Λ,存在分散控制器,使闭环极点为 Λ,则称系统可用分散控制任意配置极点. 系统可用分散控制任意匹配极点的充分必要条件是系统没有固定模.

分散系统的状态估计(state estimation for decentralized system) 根据分散量测信息对系统状态的一种估计. 设系统由 N 个子系统互联组成,即

$$\begin{cases}\dot{x}_i = A_i x_i + \sum_{j=1}^{N} H_{ij} x_j + B_i u_i, \\ y_i = C_i x_i \quad (i = 1,2,\cdots,N),\end{cases}$$

其中 x_i, y_i, u_i 分别是子系统的状态、输出和输入. (C_i, A_i) 可观测,在每个子系统建立子观测器,观测子系统状态 x_i,进而整个系统状态也可观测,由此得到的观测器称为分散系统的观测器. 分散观测器分为两大类. 一类是允许不同站间有信息交换,子观测器形式为

$$\dot{\hat{x}}_i = (A_i - K_iC_i)\hat{x}_i + K_iy_i + B_iu_i + \sum_{j=1}^{N}K_{ij}y_j$$
$$(i = 1,2,\cdots,N).$$

另一类是不允许各站间信息交换，形式为

$$\begin{cases}\dot{z}_i = F_iz_i + L_iy_i,\\ \hat{x}_i = Hz_i \quad (i = 1,2,\cdots,N).\end{cases}$$

分散观测器与集中系统有不可观测状态的观测器之间有密切联系.

分散干扰解耦(decentralized disturbance decoupling) 分散控制系统中的一种控制设计方法. 用局部控制器实现干扰解耦. 给定 N 个控制站的分散控制系统

$$\begin{cases}\dot{x} = Ax + \sum_{i=1}^{N}B_iu_i + Dv,\\ y_i = C_ix \quad (i \in \underline{N}),\\ z = Hx,\end{cases}$$

其中 x 是系统状态，u_i，y_i 分别是第 i 个控制站的输入和输出，z 是被调输出，v 是干扰输入. 所谓分散干扰解耦问题即是寻找局部控制器

$$\begin{cases}\dot{w}_i = Q_iw_i + L_iy_i \quad (i \in \underline{N}),\\ u_i = M_iw_i + K_iy_i,\end{cases}$$

使得在闭环系统中从干扰输入 v 到被调输出 z 的传递函数恒等于零.

交叠分散控制(overlapping decentralized control) 允许子系统状态有重叠部分的分散控制设计方法. 给定线性系统 Σ 和 $\widetilde{\Sigma}$：

$$\Sigma:\begin{cases}\dot{x} = Ax + Bu,\\ y = Cx,\end{cases}$$

$$\widetilde{\Sigma}:\begin{cases}\dot{\widetilde{x}} = \widetilde{A}\widetilde{x} + \widetilde{B}u,\\ y = \widetilde{C}\widetilde{x},\end{cases}$$

其中 x 和 $\widetilde{x}$ 分别是 n 维和 $\widetilde{n}$ 维状态向量，$\widetilde{n}\geqslant n$. 记 $x(t,x_0,u)$ 为系统 Σ 由初始状态 x_0 出发，在输入 $u(\tau)$，$\tau\in[0,t)$ 作用下时刻 t 的状态. 类似地，记 $\widetilde{x}(t,\widetilde{x}_0,u)$ 为系统 $\widetilde{\Sigma}$ 由初始状态 $\widetilde{x}_0$ 出发，在输入 $u(\tau)$，$\tau\in[0,t)$ 作用下时刻 t 的状态. 若存在常数矩阵 T，使对任何固定输入 $u(\tau)$，$\tau\in[0,t)$ 和 Σ 的初始状态 x_0，$\widetilde{x}_0=Tx_0$，隐含着对一切 $t\geqslant 0$，有

$$x(t,x_0,u) = T^I\widetilde{x}(t,\widetilde{x}_0,u), y[x(t)] = y[\widetilde{x}(t)],$$

其中 T^I 是 T 的广义逆，则称系统 $\widetilde{\Sigma}$ 为系统 Σ 的扩张(或 Σ 是 $\widetilde{\Sigma}$ 的收缩).

对 Σ 设计控制器时，由于所采用的信息可能存在某种约束，或计算方面的考虑，可以将其扩张为 $\widetilde{\Sigma}$，然后对 $\widetilde{\Sigma}$ 设计控制器. 在一定的条件下，为 $\widetilde{\Sigma}$ 设计的控制器在 $\widetilde{\Sigma}$ 达到的效果(如闭环稳定)可以移植到 Σ 上产生同样的效果. 尽管 $\widetilde{\Sigma}$ 的控制信息有标准的分散化形式，但它的各个控制站的子状态是 Σ 的部分状态，故当 $\widetilde{n}>n$ 时，各子状态有重叠部分，如 $x=(x_1^{\mathrm{T}},x_2^{\mathrm{T}},x_3^{\mathrm{T}})^{\mathrm{T}}$. $\widetilde{\Sigma}$ 为两个控制站，$\widetilde{x}=(\widetilde{x}_1^{\mathrm{T}},\widetilde{x}_2^{\mathrm{T}})^{\mathrm{T}}$，$\widetilde{x}_1=(x_1^{\mathrm{T}},x_2^{\mathrm{T}})^{\mathrm{T}}$，$\widetilde{x}_2=(x_2^{\mathrm{T}},x_3^{\mathrm{T}})^{\mathrm{T}}$，称这种设计方法为交叠分散控制设计.

分散自适应控制(decentralized adaptive control) 以分散控制方式实现的自适应控制. 一个分散控制系统，仅给定了系统的结构形式，某些参数未知，$I_i=\{y_i(\tau),u_i(\tau),\tau\in[0,t)\}$ 是第 i 个控制站至 t 时刻为止接收到的信息. 所谓分散自适应控制就是对每一控制站 i，仅以信息 I_i 作为输入，设计自适应控制器 Σ_i，以使整个闭环系统达到某种设计要求.

分散二次最优控制(decentralized quadratic optimal control) 分散控制系统中用分散线性状态反馈的二次最优控制方法. 设系统为

$$\dot{x}_i = \sum_{j=1}^{N}A_{ij}x_j + B_iu_i \quad (i = 1,2,\cdots,N).$$

所谓分散二次最优控制就是要找 N 个线性分散状态反馈

$$u_i = - K_ix_i \quad (i = 1,2,\cdots,N),$$

使得二次损失函数极小：

$$J = \min\int_0^{\infty}(x^{\mathrm{T}}Qx + u^{\mathrm{T}}Ru)\mathrm{d}t,$$

式中 $x=[x_1^{\mathrm{T}},x_2^{\mathrm{T}},\cdots,x_N^{\mathrm{T}}]^{\mathrm{T}}$，$u=[u_1^{\mathrm{T}},u_2^{\mathrm{T}},\cdots,u_N^{\mathrm{T}}]^{\mathrm{T}}$，$Q$ 和 R 分别是半正定阵和正定阵. 分散二次最优控制是对状态反馈阵有约束的最优控制，目前求解方法是迭代法.

广义分散控制系统(generalized decentralized control system) 兼有广义系统和分散控制系统特征的一类控制系统. 具有 N 个控制站的线性定常广义分散控制系统的一般形式是

$$E\dot{x}(t) = Ax(t) + \sum_{i=1}^{N}B_iu_i(t),$$
$$y_i(t) = C_ix(t) \quad (i \in \underline{N}),$$

其中 E 是奇异方阵，$u_i(t)$ 和 $y_i(t)$ 分别是第 i 个控制站的输入和输出，$u_i(t)$ 仅依赖于 $u_i(\tau)$，$y_i(\tau)$，$\tau\in[0,t)$.

分散正常化(decentralized normalization) 用反馈能将广义分散控制系统化为正常分散控制系统的过程. 对于广义分散控制系统 Σ：

$$\begin{cases}E\dot{x}(t) = Ax(t) + \sum_{i=1}^{N}B_iu_i(t),\\ y_i(t) = C_ix(t) \quad (i \in \underline{N}),\end{cases}$$

若存在反馈 $u_i=-K_iy_i(i\in\underline{N})$，使得

$$E + \sum_{i=1}^{N}B_iK_iC_i$$

是满秩阵，则称 Σ 是能分散正常化的. 此时，闭环系统化为正常分散控制系统.

分散无穷固定模(decentralized infinite fixed mode) 广义分散控制系统中用分散反馈无法消去

的无穷固定模. 对广义分散控制系统 Σ：

$$\begin{cases} E\dot{x}(t) = Ax(t) + \sum_{i=1}^{N} B_i u_i(t), \\ y_i(t) = C_i x(t) \quad (i \in \underline{N}), \end{cases}$$

若对任意反馈律 $u_i = -K_i y_i(t)$ $(i \in \underline{N})$，闭环系统总存在脉冲运动，则称 Σ 有分散无穷固定模.

结构等价(structure equivalence)　反映线性系统结构的一种概念. 若一个矩阵中的元素或者是固定的零元素，或者是可彼此独立变化的未定元，则称其为结构矩阵. 当一个结构矩阵中的未定元取定一组数后，所形成的常数矩阵称为此结构矩阵的一个(允许)实现. 若两个同维数的常数矩阵中零元素的位置完全相同，则可认为它们是同一个结构矩阵的两个实现，故称它们是结构等价的. 若两个线性定常系统的矩阵表示是结构等价的，则称它们为结构等价系统.

结构能控性(structural controllability)　反映系统能控性与其结构间关系的一种概念. 对线性定常系统

$$\Sigma: \quad \dot{x}(t) = Ax(t) + Bu(t),$$

若存在与系统 Σ 结构等价且在通常意义下能控的系统，则称 Σ 是结构能控的. 结构能控性反映了系统结构(它表现为矩阵 A 和 B 中零元素的位置)对其能控性的影响，而与 A 和 B 中非零元素的具体数值无关. 上述系统结构能控的充分必要条件是以下两条同时成立：

1. 不存在置换阵 P，使得

$$[PAP^{\mathrm{T}}, PB] = \begin{pmatrix} A_1 & A_2 & \vdots & B_1 \\ 0 & A_3 & \vdots & 0 \end{pmatrix},$$

式中方阵 A_3 的阶数大于零.

2. 在矩阵 $[A, B]$ 中可找出 n 个(n 为系统阶数)彼此不同行、不同列的非零元.

上述充分必要条件也常用图论方法来表达.

结构固定模(structural fixed mode)　结构等价系统在相同分散反馈下的固定模. 给定一线性定常分散控制系统：

$$\begin{cases} \dot{x}(t) = Ax(t) + \sum_{i=1}^{N} B_i u_i(t), \\ y_i(t) = C_i x(t) \quad (i = 1, 2, \cdots, N). \end{cases}$$

若每一个与此系统结构等价的系统在同样分散反馈律下都有固定模，则称此系统有结构固定模. 产生通常意义下的固定模有两方面的原因：一是系统的结构，即系统的联接方式，它决定了系统的矩阵表达式中零元素的位置；二是系统参数值间的恰当匹配. 由于实际系统的参数值很难完全精确地确定，故无法保证系统表达式中的参数匹配关系是绝对可靠的. 采用结构固定模的概念排除了由第二种原因产生的固定模，它反映了系统的结构特性对产生固定模的影响. 结构固定模的研究是受结构能控性研究成果的启发而发展起来的. 上述系统有结构固定模的充分必要条件是以下两条件之一成立：

1. 存在划分 $\{1, 2, \cdots, N\} = \{i_1, i_2, \cdots, i_S\} \cup \{i_{S+1}, \cdots, i_N\}$ 和置换阵 P，使得

$$\begin{pmatrix} PAP^{\mathrm{T}} & \vdots & PB^S \\ \cdots & & \cdots \\ C^{S^*}P^{\mathrm{T}} & \vdots & 0 \end{pmatrix} = \begin{pmatrix} A_{11} & 0 & 0 & \vdots & 0 \\ A_{21} & A_{22} & 0 & \vdots & 0 \\ A_{31} & A_{32} & A_{33} & \vdots & B_3 \\ \cdots & & & & \cdots \\ C_1 & 0 & 0 & \vdots & 0 \end{pmatrix},$$

其中

$$B^S \triangleq [B_{i_1}, \cdots, B_{i_S}],$$

$$C^{S^*} \triangleq [C_{i_{S+1}}^{\mathrm{T}}, \cdots, C_{i_N}^{\mathrm{T}}]^{\mathrm{T}},$$

A_{11}，A_{22} 和 A_{33} 均为方阵，且 A_{22} 阶数 >0.

2. 存在划分 $\{1, 2, \cdots, N\} = \{i_1, i_2, \cdots, i_S\} \cup \{i_{S+1}, \cdots, i_N\}$，使得在

$$\begin{pmatrix} A & B^S \\ C^{S^*} & 0 \end{pmatrix} \quad (B^S,\ C^{S^*}\ \text{定义同条件 1})$$

中不能选出 n 个(n 为系统阶数)彼此不同行、不同列的非零元.

上述充分必要条件也常用图论方法表示.

随机大系统(stochastic large scale system)　一种大系统. 指考虑了随机因素的大规模系统. 通常是指由随机方程(包括随机函数方程、随机微分方程、随机微分差分方程、随机积分方程、随机泛函微分方程等)描述的大系统.

互联随机系统(interconnected stochastic system)　由许多随机子系统通过耦合项连结构成的一类随机大系统. 若系统是由下面的伊藤随机微分方程描述

$$\mathrm{d}x_i(t) = A_i x_i(t)\mathrm{d}t + \sum_{j=1}^{m_i} F_{ij} x_i(t)\mathrm{d}W_{ij}(t) + \sum_{k=1}^{N} A_{ik} x_k(t)\mathrm{d}t \quad (i = 1, 2, \cdots, N), \quad (1)$$

其中 $x_i \in \mathrm{R}^{n_i}$ 是每个子系统的状态变量，A_i 和 F_{ij} 为 $n_i \times n_i$ 常矩阵，A_{ik} 为 $n_i \times n_k$ 常矩阵，$W_i = (W_{i1}, W_{i2}, \cdots, W_{im_i})^{\mathrm{T}}$，$W_i = \{W_i(t), t \geqslant 0\}$ 是一个定义在完全概率空间 $(\Omega, \mathscr{F}, P)$ 上具有独立分量的 m_i 维标准维纳过程. 互联随机系统(1)可看成是下面 N 个孤立子系统的一个线性连结：

$$\mathrm{d}x_i(t) = A_i x_i(t)\mathrm{d}t + \sum_{j=1}^{m_i} F_{ij} x_i(t)\mathrm{d}W_{ij}(t) \quad (i = 1, 2, \cdots, N).$$

复合随机系统(composite stochastic system)　即“互联随机系统”.

分布式随机迭代过程(distributed stochastic it-

erative process）一种离散时间参数随机大系统的分布式设计形式.具有离散时间参数的随机大系统通常由随机差分方程描述.由于现代科学计算的一个趋势是分布式处理,分布式处理具有响应快、可靠性高和资源共享等优点,因此许多算法被设计成分布式结构,以使能在分布式计算机系统上进行计算.分布式随机迭代过程正是离散时间参数随机大系统的分布式设计形式.如随机迭代过程

$$x(k+1)=x(k)+f(k,x(k),w(k)) \quad (1)$$

被分解为下面的分布式随机迭代过程

$$\begin{aligned}x_i(k+1)=&x_i(k)+\alpha_i(k)s_i(k,x_i(k),w_i(k))\\&+\alpha_i(k)h_i(k,x(k,w(k)))\ (i=1,2,\cdots,N),\end{aligned}$$

其中 $x(k)\in\mathrm{R}^n$ 是过程(1)在第 $k\in T=\{0,1,2,\cdots\}$ 时刻的状态;$\{w(k),k\in T\}$是一个定义在完全概率空间$(\Omega,\mathscr{F},P)$上的 m 维随机过程;$x_i(k)\in\mathrm{R}^{n_i}(n_i\leqslant n)$是第 i 个子过程在时刻 $k\in T$ 的状态,纯量序列$\{\alpha_i(k),k\in T\}$($0<\alpha_i(k)<1$)分别表示每个子过程的乘子,$w_i(k)\in\mathrm{R}^{m_i}(m_i\leqslant m)$为 $w(k)$的子向量.

具有多层递阶结构的随机大系统(stochastic large scale system with multilevel hierarchical structure）一类具层次特征的随机大系统.其构成是:第一层子系统通过某种方式的相互连结得到第二层子系统,而第二层子系统又通过某种方法的相互连结构成第三层子系统,最后构成整个随机大系统.以这种形式构成的随机大系统就称为具有多层递阶结构的随机大系统.

分散随机控制(decentralized stochastic control）有随机干扰作用下分散系统的控制方式.设系统表示为

$$x(k+1)=Ax(k)+\sum_{i=1}^{n}B_iu_i(k)+\xi(k),$$

$$y_i(k)=C_ix(k)+\eta_i(k)\ (i=1,2,\cdots,n),$$

其中 $\xi(k)$是系统输入干扰,η_i 是第 i 个站的量测干扰,它们的统计信息已知.所谓分散随机控制就是要寻找 n 个分散控制器

$$\begin{aligned}u_i(k)=f_i(&y_i(1),y_i(y_i(2)),\cdots,y_i(k),\\&u_i(1),u_i(2),\cdots,u_i(k-1))\\&(i=1,2,\cdots,n),\end{aligned}$$

使目标函数

$$J=E\left\{\frac{1}{2}\sum_{j=1}^{k}[x^{\mathrm{T}}(j)Q(j)x(j)+\sum_{i=1}^{n}u_i^{\mathrm{T}}(j)R_i(j)u_i(j)]\right\}$$

达到极小,其中 $Q(j)$,$R(j)$是正定阵.非线性或连续时间分散随机控制可以类似定义.

随机大系统的分散镇定(decentralized stabilization of stochastic large scale system）用分散状态反馈或输出反馈,使受控的随机大系统在适当的稳定性意义下渐近稳定的控制方法.这就是随机大系统的镇定.由于随机大系统在许多情形下具有信息分散化的特点,因此人们自然就希望利用分散的信息实现分散化的镇定.当随机大系统被划分为许多随机子系统时,每个随机子系统分别由独立的分散控制器进行控制,每个分散控制器只观测系统的局部信息,并以此构成局部控制律,使随机大系统镇定.这就是随机大系统的分散镇定.

随机大系统的p阶矩稳定性(p-moment stability of stochastic large scale system）随机大系统的一种特殊稳定性.设 $x(t)=x(t;t_0,x_0)$是随机大系统具有初始条件(t_0,x_0)的任一解.如果对于所有满足 $p_1+p_2+\cdots=p$ 的非负整数 $p_1,p_2,\cdots$,对任给的正数 $\varepsilon>0$,存在正数 $\delta>0$ 使得当$|x_0|<\delta$ 时,成立

$$|E[x_1^{p_1}(t)x_2^{p_2}(t)\cdots]|<\varepsilon\ (t\geqslant t_0),$$

其中 x_i 是 x 的第 i 个分量,则称随机大系统是 p 阶矩稳定的.

随机大系统的p阶矩全局指数稳定性(p-moment globally exponential stability of stochastic large scale system）随机大系统的一种特殊的全局指数稳定性.设 $x(t)=x(t;t_0,x_0)$是随机大系统具有初始条件(t_0,x_0)的解.如果对于所有满足 $p_1+p_2+\cdots=p$ 的非负整数 $p_1,p_2,\cdots$,它是 p 阶矩稳定的,并且对于任给的 $x_0\in\mathrm{R}^n$,存在正数 $\alpha,\beta>0$,使得下式成立:

$$|E[x_1^{p_1}(t)x_2^{p_2}(t)\cdots]|\leqslant\alpha\exp\{-\beta(t-t_0)\}\quad(t\geqslant t_0),$$

其中 $x_i(t)$是向量 $x(t)$的第 i 个分量,则称随机大系统是 p 阶矩全局指数稳定的.

滞后随机大系统(delay stochastic large scale system）带有时间滞后的一类随机大系统.对于随机大系统,信号的传递不可避免地产生滞后,在某些情形下滞后对随机大系统的影响可以忽略,而在许多情形下滞后的影响就必须给予考虑,这就得到了带有滞后的随机大系统——滞后随机大系统.由伊藤随机微分差分方程描述的滞后随机大系统可表示为

$$\begin{aligned}\mathrm{d}x_i(t)=&A_ix_i(t)\mathrm{d}t+\sum_{j=1}^{m_i}F_{ij}x_i(t)\mathrm{d}w_{ij}(t)\\&+\sum_{k=1}^{N}A_{ik}x_k(t-\tau_{ik})\mathrm{d}t\ (i=1,2,\cdots,N),\end{aligned}$$

其中 $x_i\in\mathrm{R}^{n_i}$是第 i 个子系统的状态变量,A_i 是一个 $n_i\times n_i$ 常矩阵,$w_i=(w_{i1},w_{i2},\cdots,w_{im_i})^{\mathrm{T}}$ 是一个 m_i 维具有独立分量的零均值标准维纳过程,它表示作用在第 i 个子系统上的随机扰动,F_{ij}是一个 $n_i\times n_k$ 互

联矩阵,τ_{ik}是一个非负实数.上面各个指标的取值为 $i,k=1,2,\cdots,N$.

滞后随机大系统的滞后无关均方渐近稳定性(delay-independent mean-square asymptotic stability of delay stochastic large scale system) 滞后随机大系统独立于滞后的一种稳定性.如果对于滞后随机大系统中的任意非负实数 $r_i(i=1,2,\cdots,N)$,系统的平衡态 $x(t)\equiv 0$ 总是保持均方渐近稳定的,那么称滞后随机大系统是滞后无关均方渐近稳定的.

多滞后连续系统的比较原理(comparison principle for continuous system with multi-delay) 适用于多滞后连续系统的比较原理.多滞后线性定常连续微分不等式系统

$$\dot{x}_i(t)\leqslant\sum_{j=1}^{n}a_{ij}^{(0)}x_j(t)+\sum_{r=1}^{N}\sum_{j=1}^{n}a_{ij}^{(r)}x_j(t-\tau_{ij}^{(r)})\quad(i=1,2,\cdots,n).\tag{1}$$

记 $\tau=\max[\tau_{ij}^{(r)}>0,i,j=1,2,\cdots,n,r=1,2,\cdots,N]$.在区间 $t_0-\tau\leqslant t\leqslant t_0$ 上给定初值函数 $x_{0j}(t),x_j(t)=x_{0j}(t),t_0-\tau\leqslant t\leqslant t_0(j=1,2,\cdots,n)$.建立(1)的辅助比较系统

$$\dot{y}_i(t)=\sum_{j=1}^{n}a_{ij}^{(0)}y_j(t)+\sum_{r=1}^{N}\sum_{j=1}^{n}a_{ij}^{(r)}y_j(t-\tau_{ij}^{(r)}),\tag{2}$$

在区间 $t_0-\tau\leqslant t\leqslant t_0$ 上给定初值函数 $y_{0j}(t),y_j(t)=y_{0j}(t),t_0-\tau\leqslant t\leqslant t_0(j=1,2,\cdots,n)$.比较原理是:当系数 $a_{ij}^{(0)},a_{ij}^{(r)}$满足下列不等式:

$$\begin{cases}a_{ii}^{(0)}<0,\ i=1,2,\cdots,n;\\ a_{ij}^{(0)}\geqslant 0,\ i\neq j\quad(i,j=1,2,\cdots,n);\\ a_{ij}^{(r)}\geqslant 0,\ i,j=1,2,\cdots,n,\ r=1,2,\cdots,N\end{cases}$$

时,由初值函数满足的不等式

$$x_{0j}(t)\leqslant y_{0j}(t),\ t_0-\tau\leqslant t\leqslant t_0\quad(j=1,2,\cdots,n),$$

能导出下列不等式:

$$x_j(t)\leqslant y_j(t)$$
$$(j=1,2,\cdots,n,\ t_0-\tau\leqslant t\leqslant+\infty),$$

式中,$x_j(t)$和 $y_j(t)$分别是多滞后系统(1)和辅助比较系统(2)的解.在一定条件下,多滞后时变线性系统或多滞后非线性系统的比较原理也成立.

多实数滞后离散系统的比较原理(comparison principle for discrete-time system with real multi-delay) 适用于多实数滞后离散系统的比较原理.多实数滞后线性定常离散系统

$$x_i(k+1)\leqslant\sum_{j=1}^{n}a_{ij}^{(0)}x_j(k)+\sum_{r=1}^{N}\sum_{j=1}^{n}a_{ij}^{(r)}x_j(k-\tau_{ij}^{(r)})\quad(i=1,2,\cdots,n).\tag{1}$$

令 $\tau=\max[\tau_{ij}^{(r)}>0,i,j=1,2,\cdots,n,r=1,2,\cdots,N]$,在区间$-\tau\leqslant k\leqslant 0$ 上给定初值函数 $x_{0j}(k)$,且 $x_j(k)=x_{0j}(k),j=1,2,\cdots,n,-\tau\leqslant k\leqslant 0$.(1)的辅助比较系统为

$$y_i(k+1)=\sum_{j=1}^{n}a_{ij}^{(0)}y_j(k)+\sum_{r=1}^{N}\sum_{j=1}^{n}a_{ij}^{(r)}y_j(k-\tau_{ij}^{(r)})\quad(i=1,2,\cdots,n).\tag{2}$$

并在区间$-\tau\leqslant k\leqslant 0$ 上给定初值函数 $y_{0j}(k)$,且

$$y_j(k)=y_{0j}(k)\ (j=1,2,\cdots,n,\ -\tau\leqslant k\leqslant 0).$$

离散系统的比较原理为:当系数 $a_{ij}^{(0)},a_{ij}^{(r)}$满足不等式 $a_{ij}^{(r)}\geqslant 0(i,j=1,2,\cdots,n,r=1,2,\cdots,N)$时,由初始函数满足的不等式

$$x_{0j}(k)\leqslant y_{0j}(k)\ (-\tau\leqslant k\leqslant 0,\ j=1,2,\cdots,n),$$

能导出如下不等式:

$$x_j(k)\leqslant y_j(k)\quad(j=1,2,\cdots,n,\ k=0,1,2,\cdots),$$

式中 $x_j(k)$和 $y_j(k)(j=1,2,\cdots,n)$分别是多实数滞后线性定常离散系统(1)及辅助比较系统(2)的解.一定条件下,多实数滞后时变离散系统和多实数滞后非线性离散系统的比较原理成立.

无滞后系统与滞后系统的稳定性等价(stability equivalence between non-delay system and delay system) 用无滞后系统研究滞后系统稳定性的一种方法.指用无滞后的线性系统

$$\dot{x}_i(t)=\sum_{j=1}^{n}(a_{ij}+b_{ij})x_j(t)\quad(i=1,2,\cdots,n),\tag{1}$$

或无滞后的非线性系统

$$\dot{x}_i(t)=\sum_{j=1}^{n}(a_{ij}+b_{ij})x_j(t)+F_i(x(t),x(t))\quad(i=1,2,\cdots,n)\tag{1$_1$}$$

的零解的稳定性代替具有滞后的线性系统

$$\dot{x}_i(t)=\sum_{j=1}^{n}a_{ij}x_j(t)+\sum_{j=1}^{n}b_{ij}x_j(t-\tau_{ij})\quad(i=1,2,\cdots,n),\tag{2}$$

或具有滞后的非线性系统

$$\dot{x}_i(t)=\sum_{j=1}^{n}a_{ij}x_j(t)+\sum_{j=1}^{n}b_{ij}x_j(t-\tau_{ij})+F_i(x(t),x(t-\tau))\quad(i=1,2,\cdots,n)\tag{2$_1$}$$

的零解的稳定性,这里

$$F(x(t),x(t-\tau))=\sum_{\sum_{j=1}^{n}(l_j+m_j)\geqslant 2}P_i^{(l_1,\cdots,l_n,m_1,\cdots,m_n)}\cdot(x_1^{l_1}(t),\cdots,x_n^{l_n}(t),x_1^{m_1}(t-\tau_{i1}),\cdots,x_n^{m_n}(t-\tau_{in})),$$

并且有正数 ε,使

$$\sum_{\sum_{j=1}^{n}(l_j+m_j)\geqslant 2}|P_i^{(l_1,\cdots,l_n,m_1,\cdots,m_n)}|\varepsilon^{l_1+\cdots+l_n+m_1+\cdots+m_n}<\infty.$$

控制系统与滞后控制系统的镇定等价(stabilized equivalence between control system and delay control system) 研究滞后系统稳定性的一种方法.用线性(或非线性)控制系统

$$\dot{x}_i = \sum_{j=1}^{n} (a_{ij}^{(1)} + a_{ij}^{(2)}) x_j + \sum_{j=1}^{n} (b_{ij}^{(1)} + b_{ij}^{(1)}) u_j \quad (i = 1,2,\cdots,n) \quad (1)$$

的闭环系统的渐近稳定性代替滞后线性(或非线性)控制系统

$$\dot{x}_j(t) = \sum_{j=1}^{n} a_{ij}^{(1)} x_j(t) + \sum_{j=1}^{n} a_{ij}^{(2)} x_j(t - \tau_{ij}^{(1)}) + \sum_{j=1}^{n} b_{ij}^{(1)} u_j + \sum_{j=1}^{n} b_{ij}^{(2)} u_j \quad (2)$$

的闭环系统的渐近稳定性,称为控制系统(1)与滞后控制系统(2)的镇定等价.

微 分 对 策

微分对策(differential game)　一种考虑了动态特性的对策.它首先是由军事上的需要引起的,以后又进入工程、经济等领域.

20 世纪 40 年代中期,美国兰德公司(Rand Corporation)受到了对策理论的强烈冲击,这是因为当时出版了由冯·诺伊曼(von Neumann,J.)和莫根施特恩(Morgensten,O.)合写的第一本对策理论的书《Theory of Games and Economic Behavior》(Princeton University Press,1944),兰德公司关心军事问题,既然对策理论是分析对抗的,对抗分析又是军事筹划的方法,因此有理由将对策理论应用于军事筹划.于是在空军的赞助下,该公司组织以埃萨克斯(Isaacs,R.)为首,展开了这方面的研究工作.

开头的研究是关于导弹拦截飞行器的,他们发表的第一篇论文是“追击对策”(Isaacs,R.,Games of Pursuit,Rand Report,p257,17,Nov. 1951).但一段时间后进展不大,发觉用已有的对策理论解决拦截问题不能奏效,于是另辟蹊径,寻找新方法.以后,埃萨克斯独立提出了实质上是贝尔曼(Bellman,R.)的动态规划理论中的基本方程(所以又称贝尔曼-埃萨克斯方程),并结合各种奇异曲面做重点分析,使研究取得了一系列进展.

事实上,微分对策的形成和发展有两个根源:一是对策理论,一是控制理论,特别是最优控制理论.以导弹拦截飞机为例,参与者是两方,且均为机动、可控.一方追击,一方躲避,双方对抗,并各自选取对自已最有利的控制方案.既然有双方对抗的性质,所以和对策论相关,那里的一些概念,诸如策略、对策值、两人零和、鞍点等都将被借用.但对策理论所处理的问题与微分对策相比,在某种意义上可看做是静态的,而微分对策是动态的,因此,单是用对策理论的方法已不能解决问题.

另一方面,控制理论已在飞速发展,以与微分对策密切相关的最优控制理论来看,两者之间的关系可认为是由单方控制发展到双方或多方控制,而且增添了对抗或合作(如两卫星交会问题)的因素.最优控制也研究动态问题,而且它和两人零和性质的微分对策都是考虑极值问题,因而更有理由期望借用和发展其中的方法,动态规划正是最优控制理论中的一种重要方法,因此贝尔曼和埃萨克斯各自研究互不相同、但却密切相关的问题,于是各自独立提出本质相同的方法不是偶然的.

埃萨克斯的主要工作是 1954 年至 1955 年的四篇兰德公司的内部报告,以后于 1965 年公开,并整理成书《Differential Games》(John Wiley and Sons Inc.,1965).这是第一部关于微分对策理论的专著,“微分对策”一词也是这时提出的.此书的出版标志着开创了微分对策这一新的理论.

微分对策问题(topic of differential games)　双方(多方)对抗或合作的动态对策问题.由于对动态的描写是用微分方程,故称微分对策.在一类最基本的微分对策(定量微分对策)中,对策双方都欲使目标函数最优,一方使最大,一方使最小,故实际上是一种双方最优控制问题.微分对策的产生可看做来自两个方面的发展,一是自对策论由静态发展到动态,一是自最优控制理论由单方控制发展到双方控制.微分对策理论是由埃萨克斯(Isaacs,R.)建立的,并于 1965 年出版了这一理论的第一本书《微分对策》.

动态两人零和对策(dynamic two-person zero-sum game)　一种微分对策.指微分对策问题中对抗双方得失相抵,其和为零的对策.在定量微分对策中,参与对策的双方具有相同形式的目标函数(性能指标),一方要选取其策略,使此函数取最小可能的值,而另一方则选取其策略,使此函数取最大可能的值,双方对抗,且一方有所得,另一方必有所失,其量相同,即得失相抵,其和为零.这类问题称为动态两人零和对策.

定量微分对策(quantitative differential games or games of degree)　一种微分对策.定量微分对策是双方对于某一量(性能指标)的极值的竞争.问题如下:对策的动态过程为

$$\dot{x} = f(x,u,v,t),\ x(t_0) = x_0\ (u \in U, v \in V),$$

其中状态 x 为 n 维,t_0 为初始时间,u,v 分别为 A,B 两方的策略,各为 r 及 s 维,U,V 为策略集,可为有界闭集.性能指标及终端约束分别为

$$J(u,v) = \varphi(x(t_1),t_1) + \int_{t_0}^{t_1} L(x,u,v,t)\mathrm{d}t,$$

$$\psi(x(t_1),t_1) = 0,$$

这里 t_1 为终止时间.A,B 两方要从策略集中选($\bar{u}$,

$\bar{v}$)，使 $J(\bar{u},v)\leqslant J(\bar{u},\bar{v})\leqslant J(u,\bar{v})$，这样的($\bar{u},\bar{v}$)分别称为 A,B 的最优策略. 定量微分对策是最优控制从单方向双方的扩展. 定量微分对策由两方参与，竞争性能指标的极值. 最优控制问题只由一方参加，选其控制使性能指标取极值. 见"最优控制"条目，两者相比，可知最优控制是定是量微分对策由双方蜕化为单方的特例.

双方极值原理(two-sided extremum principle) 定量微分对策的基本定理. 对于定量微分对策的求解，是要找出双方的最优策略($\bar{u},\bar{v}$). 对此，有如下原理.

双方极值原理：($\bar{u},\bar{v}$)最优性的必要条件是：$\bar{u}(t),\bar{v}(t)$对于一切 $t\in[t_0,t_1]$，满足

$$\begin{aligned}&\max_{u\in U}\min_{v\in V}H(\lambda,\bar{x},u,v,t)\\&=\min_{v\in V}\max_{u\in U}H(\lambda,\bar{x},u,v,t)\\&=H(\lambda,\bar{x},\bar{u},\bar{v},t),\end{aligned}$$

其中 $\bar{x}$ 是与($\bar{u},\bar{v}$)相对应的最优轨线，$H(\lambda,x,u,v,t)$称为哈密顿函数，定义为

$$H(\lambda,x,u,v,t)\triangleq\sum_{i=1}^{n}\lambda_i f_i(x,u,v,t)-L,$$

λ_i 称伴随变量，满足

$$\dot{\lambda}_i=-\frac{\partial H}{\partial x_i}\qquad(i=1,2,\cdots,n),$$

$$\lambda_i(t_1)=-\left.\frac{\partial\psi}{\partial x_i}\right|_{t_1}\quad(i=1,2,\cdots,n),$$

$$\varphi=\varphi+\nu\psi,$$

f_i,L,φ,ψ 等参见"定量微分对策"，ν 为一正的常数乘子.

时间指标的微分对策问题(problem of differential game with time-performance) 以时间为性能指标的微分对策问题. 这类问题的一般表述为：A,B 双方的运动方程及策略为

$$\dot{x}=f(x,u,v,t),x(t_0)=x_0\ (u\in U,v\in V),$$

并于 $t=T$ 时轨线到达超曲面 $\psi(x)=0$，即$\psi(x(T))=0$. 时间 T 将取为性能指标. A,B 将分别自 U,V 中选取其最优策略($\bar{u},\bar{v}$)，使

$$T(\bar{u},v)\leqslant T(\bar{u},\bar{v})\leqslant T(u,\bar{v}).$$

两架飞机进行空战格斗，当一追一逃时，便是这类问题的一例.

线性微分对策的充分必要条件(necessary and sufficient condition for linear differential game) 在线性情况下，双方极值原理为策略最优性的充分必要条件. 若 A,B 两方的运动方程为线性，

$$\dot{x}=Ax+B(u,v),\ u\in U,v\in V,$$

且性能指标也为线性，

$$J(u,v)=\sum_{i=1}^{n}C_i x_i(t_1),$$

式中 C_i 为常数，t_1 为对策终止时间. 又终端约束 $\psi(x(t_1))=0$，其中 $\psi(x)$也为线性函数，或没有终端约束. 这时，对于满足下式的最优策略

$$J(\bar{u},v)\leqslant J(\bar{u},\bar{v})\leqslant J(u,\bar{v})$$

有如下双方极值原理：($\bar{u},\bar{v}$)最优性的充分必要条件是$u(t),v(t)$对于一切$t\in[t_0,t_1]$满足

$$\begin{aligned}&\max_{u\in U}\min_{v\in V}H(\lambda,\bar{x},u,v,t)\\&=\min_{v\in V}\max_{u\in U}H(\lambda,\bar{x},u,v,t)\\&=H(\lambda,\bar{x},\bar{u},\bar{v},t),\end{aligned}$$

其中 H,λ 等的意义参见"双方极值原理".

定性微分对策(qualitative differential game) 双方为实现某一结局的竞争对策. 它是微分对策中的另一基本问题，和定量微分对策相比，它不研究某一性能指标的极值，而是研究某种结局能否实现. 设 A,B 两方的相对运动方程为

$$\dot{x}=f(x,u,v),\ u\in U,\ v\in V,$$

给定一目标集 $\psi(x)\leqslant 0$，若满足此条件，则目标实现，对策结束. A 方要选取其策略 u，力图实现此条件，B 方则选 v，力图不实现此条件. 定性微分对策是能控制问题从单方向双方的扩展. 定性微分对策由双方参与，各自选取控制，进行到达目标的竞争. 能控制问题只是一方参与，研究选取控制，使初始状态能否达到目标集. 见"能控性"条目，两者相比，可知能控性是定性微分对策的特例. 从定性微分对策的极值性质(参见"定性微分对策的极值性质")可知，定性微分对策可化为两个极值问题. 定量微分对策也是两个极值问题(参见"定量微分对策")，即

$$J(\bar{u},v)\leqslant J(\bar{u},\bar{v})$$

$$J(\bar{u},\bar{v})\leqslant J(\bar{u},\bar{v})$$

这样，就可将这两类微分对策问题统一于它们的极值性质，从而可用统一的方法处理，得出同一形式的结论，即最优策略所应满足的必要条件都以某种极值原理的形式表达，都用哈密顿函数的极值描述，当然，其中有相应的区别.

追逃对策(pursuit-evasion game) 双方对追及与否的对策. 一类典型的定性微分对策. 参与对策的追、逃两方，分别记为 P 和 E，P 方欲追近 E，E 方相反，欲远离 P. 若 P 方接近到 E 方的一定范围，并用 $\psi(x)\leqslant 0$(目标集)表示，则称"追及"或"捕获"，这时对策结束.

界栅(barrier) 捕捉区与逃避区的分界面. 是定性微分对策问题中的一种分界面. 以追逃对策为例，若双方性能相差悬殊，则或者必能追及，或者必不能追及，且这种策略(u,v)可有无穷多，故答案不确定. 只是当双方性能处于一定均势时，才可能有明确答案. 这时状态空间中可能存在"捕捉区"和"逃避区"两个区域. 若逃方 E 位于捕捉区，则不论 E 如何

使用其策略 $v\in V$,追方 P 总可用一适当策略 $u\in U$,将 E 追及.若 E 方位于逃避区,则不论 P 方如何使用 $u\in U$,E 方总可用一适当策略 $v\in V$,使不被追及.这两个区域的分界面,称为界栅.

定性双方极值原理(qualitative two-sided extremum principle) 定性微分对策的基本定理.对策双方采取最优策略所需满足的一个原理.以追逃对策为例,若存在界栅,则当逃方 E 位于界栅上时,双方将进行最激烈的对抗,亦即均使用其最优策略 $(\bar{u},\bar{v})$.若追方 P 不使用 $\bar{u}$,则 E 可进入逃避区,若 E 不使用 $\bar{v}$,则 E 可被驱入捕捉区.当双方均作用 $(\bar{u},\bar{v})$ 时,E 将在界栅上运动,因此,若求得了 $(\bar{u},\bar{v})$ 也就找到了界栅.寻求 $(\bar{u},\bar{v})$ 可据下面的定性双方极值原理:$(\bar{u},\bar{v})$ 最优的必要条件是,$\bar{u}(t)$,$\bar{v}(t)$ 对于一切 $t\in(t_0,t_1)$ 满足

$$\begin{aligned}&\max_{u\in U}\min_{v\in V}H(\lambda,\bar{x},u,v,t)\\ =&\min_{v\in V}\max_{u\in U}H(\lambda,\bar{x},u,v,t)\\ =&H(\lambda,\bar{x},\bar{u},\bar{v},t)\equiv 0,\end{aligned}$$

这里

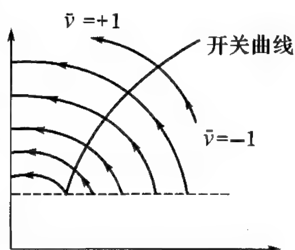
$$H(\lambda,x,u,v,t)\triangleq\sum_{i=1}^{n}\lambda_i f_i(x,u,v,t),$$

$$\dot{\lambda}_i=-\frac{\partial H}{\partial x_i}\qquad(i=1,2,\cdots,n),$$

$$\lambda_i(t_1)=-\nu\left.\frac{\partial\psi}{\partial x_i}\right|_{t_1}\quad(i=1,2,\cdots,n),$$

t_1 是自 $x(t_0)=x_0$ 出发沿界栅到达目标集边界的时间,$\psi(x)$ 是目标集的函数,ν 是一正的常数乘子.

开关曲面(switching surface) 亦称转换曲面.一种曲面.指最优策略 $(\bar{u},\bar{v})$ 的值发生跳变的曲面.例如,在某一具体问题中(快慢路径问题),当最优轨线经过开关曲面时,$\bar{v}$ 的值由 -1 跳变为 $+1$,见右图.也有这样的例子,某一开关曲面处,$\bar{u}$ 发生跳变,在另一开关曲面上,$\bar{v}$ 发生跳变等.

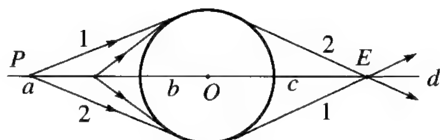


等值策略曲面(equivocal surface) 一种奇异曲面.所谓"等值策略",是指在这种曲面的某一点上,某一方可采取两种最优策略,并导致相同的对策值.

奇异曲面(singular surface) 微分对策的概念之一.指不同性质区域的界面.微分对策问题状态空间中不同性质定义的区域间的界面.在微分对策中,对于不同的问题存在有多种不同形式的奇异曲面,它们常在解题中起着支配性的作用.例如,有的奇异曲面将状态空间分为一些不同性质的区域,在每一区域中轨线的性质不同;又如,有些特殊性质的轨线,就在某种奇异曲面上运动;再如,某种奇异曲面是具有某种性质的轨线的出发地等.对于各种奇异曲面没有统一的理论,只能分别进行分析.

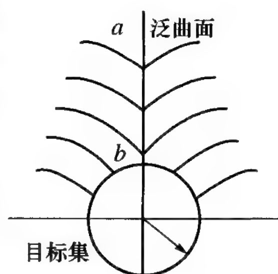
转换曲面(transition surface) 即"开关曲面".

散射曲面(dispersal surface) 一种曲面.指最优轨线的出发曲面.例如,P 追 E,P,E 间有一圆形



障碍物(见图),P,E 的初始位置将分别位于 ab 及 cd 上.P,E 的最优轨线将均为沿圆的切线.这样 ab 将为 P 的最优轨线的出发曲面,即最优轨线将自直线 ab 散射出去,故称散射曲面.

泛曲面(universal surface) 一种曲面.指最优轨线走向的曲面.在某些问题中,最优轨线将流向某一曲面,然后再沿此曲面走向目标集,这样的曲面称为泛曲面,如右图中的 ab.



哈密顿-雅可比-贝尔曼-埃萨克斯方程(Hamilton-Jacobi-Bellman-Isaacs equation) 微分对策问题中一种基本方程,即形如下式的方程:

$$-\frac{\partial V_i}{\partial t}=\min_{u_i\in U_i}\left[\frac{\partial V_i}{\partial t}f(t,x,u)+g(t,x,u)\right]$$
$$(i=1,2,\cdots,N),$$

其中 $u=(u_1,u_2,\cdots,u_N)$,u_i 为决策者 P_i 的决策值.此方程是由如下的对策系统产生的:

$$\dot{x}=f(t,x,u),\ x(0)=x_0,\ t\geqslant 0,$$

$$J_i=\int_0^T g(t,x,u)\mathrm{d}t+q(T,x(T))$$
$$(i=1,2,\cdots,N).$$

而方程中的 V_i 定义为

$$V_i(t,x)=\min_{u_i(s)}\left[\int_t^T g(s,x(s),u(s))\mathrm{d}s+q(T,x(T))\right].$$

由于此方程曾在最优控制、动态规划中应用,以后又由埃萨克斯(Isaacs,R.)应用到微分对策中来,故称之为哈密顿-雅可比-贝尔曼-埃萨克斯方程.

目标集(termination set) 一种集合.是定性微分对策结束时所要求实现的集合.在定性微分对策中,对策结束时所要求实现的条件界定的集合.如用

导弹拦截飞机，要求在拦截过程结束时，飞机处于导弹爆炸时能击毁飞机的有效范围内，亦即导弹与飞机间的距离不超过某一给定值. 一般地，目标集可写成 $\psi(x)\leqslant 0$.

定性微分对策的极值性质(extremum properties of qualitative differential game) 定性微分对策也是一种极值控制. 定性微分对策可归结为能否将某一初始状态控制到一个给定目标集 $\psi(x)\leqslant 0$ 的问题. 在定性极值原理的证明中，论证双方各自的最优控制$(\bar{u},\bar{v})$所应满足的条件时，对于 $\bar{u}$，是从下面的关系式出发：

$$\begin{cases}(\bar{u},\bar{v}),\psi(\bar{x}(t_1))=0,\\(u,\bar{v}),\psi(x_{\bar{v}}(t_1))>0,\end{cases}\tag{1}$$

其中，$\bar{x}$ 是与$(\bar{u},\bar{v})$相对应的最优轨线，$x_{\bar{v}}$是与$(u,\bar{v})$相对应的轨线，t_1 是对策结束的时间. 对于 $\bar{v}$，相应地有

$$\begin{cases}(\bar{u},\bar{v}),\psi(\bar{x}(t_1))=0,\\(\bar{u},v),\psi(x_{\bar{u}}(t_1))<0,\end{cases}\tag{2}$$

这实质上是两个极值问题. 式(1)说明，与轨线 $x_{\bar{v}}$相比，沿 $\bar{x}$ 在 t_1 时将给出 $\psi(x)$以极小值. 式(2)说明，与 $x_{\bar{u}}$相比，沿 $\bar{x}$ 在 t_1 时将给出 $\psi(\bar{x})$以极大值. 从而将定性微分对策问题化为解两个极值问题. 这就是定性微分对策问题的极值性质.

全向推力火箭对策(isotropic rocket game) 定量微分对策中一典型对策. 它是两火箭追与躲之间的对策. 这是定量微分对策的一个典型例子. 用火箭 P 追击目标 E，火箭受可变方向的推力 F 作用，且此方向可控制. 双方各有一定的速度，E 的速度的方向也可控制. 设追击在平面上进行，可列出 P，E 运动的方程式. 当 P，E 间的距离不大于给定值 l 时，认为实现了"捕获". 取实现捕获的时间 T 为性能指标. 双方各选取自己的控制，P 欲使 T 最小，E 欲使 T 最大.

人车对策(murderer chauffeur game) 定性微分对策的一种. 它是步兵与战车之间的对策. 例如步兵和战车之间的战斗，战车欲碰撞步兵，步兵则躲避战车，战车速度快，但转弯不灵活，步兵速度慢，但转弯灵活，双方各有优势. 据二者的速度、转弯半径等参数，可列出在地面上二者间相对运动的微分方程，以及如下形式的目标集：

$$x^2+y^2-l^2=0.$$

这是定性微分对策的一个典型例子，在此例中，将出现泛曲面、界栅、散射曲面、等值曲面等奇异曲面.

二车对策(game of two cars) 一种微分对策. 指两战车之间的对策. 在人车对策中，将"人"的一方换为另一车，即为二车对策. 在此例中，追、逃两方各有自己的最大速度及最小转弯半径. 在此例中，将出现散射曲面和泛曲面.

海上雷达监视问题(radar surveillance-evasion problem) 一类微分对策问题. 它研究一方如何摆脱另一方的雷达监视. 海面上有船 P 和 E，各有自己的最大航速及最小转弯半径. P 上装有雷达，其有效搜索半径为 l. P 欲跟踪监视 E，E 欲摆脱监视，当 E 跑出 P 的监视半径时对策结束. 在这个问题中，目标集可写成

$$x^2+y^2-l^2\geqslant 0.$$

碰撞避免问题(collision avoidance problem) 一类合作型微分对策问题. 它研究两船(或飞机)如何合作以避免碰撞. 海上大型船舶，由于质量巨大，操纵不灵活，往往发生两船之间的碰撞事故. 空中高速飞行的飞机，也有类似情况. 为避免事故，要求双方各自操纵，使互相接近时二者间的距离不小于某一给定值. 这是一种要求特定目标集的定性微分对策问题.

点捕获问题(point-capture problem) 脱靶量为零的拦截问题. 在拦截问题中，若要求在拦截过程结束时，追逃两方完全重合，亦即二者间的距离为零，则称点捕获问题. 例如用导弹拦截导弹，要求完全命中，脱靶量为零，即属此类. 在此问题中，目标集是一个点.

空战格斗中角色二重性问题(ambiguity problem in air combat) 空战中"追躲"两方的角色互易问题. 两架飞机进行空战时，每架飞机均力图使对方进入自己前方的火力区，这样，追、逃两方的地位，就可能发生变化，追方可能变为逃方，逃方也可能变为追方，双方的"角色"不固定，可以互换. 这就是角色二重性问题. 这是一种新型的微分对策问题，需要发展新的求解方法.

有限时间局部捕捉区(partial capture region on finite time interval) 微分对策术语. 指在有限时间内捕捉区的范围. 在原有定性微分对策理论中，用"界栅"将整个状态空间划分为捕捉区及逃避区(参见"界栅")，同时对策进行的时间也不受限制. 但这种作法对于解决某些实际问题有缺点. 其一，有些问题并不需要考虑整个状态空间，也不能不考虑时间. 例如，在空战中只是在一定时间内和部分空间中才有实际意义. 再如，飞行器燃料有限，从而飞行时间和距离也都有限. 其二，界栅可能出现半途中止的现象，这时整个状态空间都是捕捉区. 但由于对方的位置不同，捕捉时间的长短仍有区分. 在某些问题中，捕捉时间长短是有意义的. 基于以上原因，乃提出"有限时间局部捕捉区"的概念. 有限时间局部捕捉区是由界栅、目标集的部分边界(可用部分)及到达目标集的最优轨线不同时间的等时线(或面)为边界所构成的区域，此区域随时间而变，这里的最优轨线

是以时间为最优指标的. 例如,在空战格斗中,某种性能的飞机其火力区前部的有限时间局部捕捉区如图 1 所示. 当敌机处于例如 40s 区域内时,则可在 40s 内使其进入火力区. 如果考虑相对运动,还可得出飞机被追击时其后部的局部危险区,如图 2 所示.

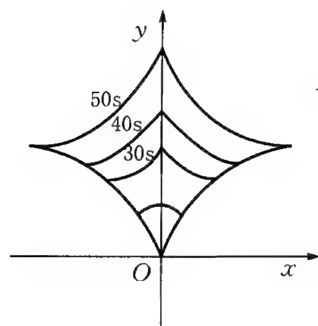


图 1

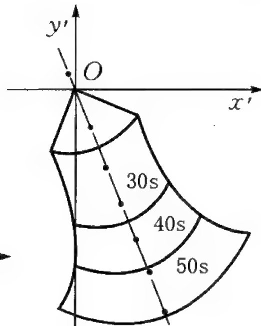


图 2

这样,在空战中,当追击敌机时,前方有一局部捕捉区,当被追击时,后方有一局部危险区,从而给出解角色二重性问题的一种方法.

目标集的可用部分(useable part of termination set) 微分对策术语. 指对策结束时目标集上实际可达到的部分. 在具有目标集的问题中,许多情况下,并不是目标集的每一处都可作为对策结束的点,而是只能到达目标集的某一部分,这称为目标集的可用部分.

确定能控区边界的方法(method of determining the boundary of controllable region) 用于确定能控区边界的一些方法. 主要有:

1. 倒转时间方法. 据定性极值原理,对于满足极大值条件及 $H\equiv 0$ 的最优轨线,若倒转时间 $t\to -\infty$ 时轨线以一个封闭的曲线为极限,则由这样的曲线所构成的封闭曲面就是能控区边界.

2. 代数方法. 对于线性定常控制系统

$$\dot{x} = Ax + Bu,$$

应用定性极值原理,只需找出满足

$$\lambda^{T}[B, AB, \cdots, A^{n-1}B] \equiv 0$$

的 λ 的正交基便可直接确定能控区.

动态非合作对策(dynamic non-cooperative game) 一类微分对策问题. 指两人或两人以上并有各自支付函数的对策. 此类对策是两人二和对策的一个推广. 对策中可以有两个或两个以上的决策者,各自的支付函数依赖于每个决策者所采取的决策. 在决策者之间期望达到纳什平衡. 这一概念是由纳什(Nash,N.)于 1951 年引入到 N 人对策中来的,故以纳什命名. 类似于鞍点的概念,纳什平衡点 $\{\gamma_1^*, \gamma_2^*, \cdots, \gamma_N^*\}$ 满足如下不等式:

$$J_1^* \triangleq J_1(\gamma_1^*, \gamma_2^*, \cdots, \gamma_N^*) \leqslant J_1(\gamma_1, \gamma_2^*, \cdots, \gamma_N^*),$$
$$J_2^* \triangleq J_2(\gamma_1^*, \gamma_2^*, \cdots, \gamma_N^*) \leqslant J_2(\gamma_1^*, \gamma_2, \cdots, \gamma_N^*),$$
$$\vdots$$
$$J_N^* \triangleq J_N(\gamma_1^*, \cdots, \gamma_{N-1}^*, \gamma_N^*) \leqslant J_N(\gamma_1^*, \cdots, \gamma_{N-1}^*, \gamma_N),$$

其中,J_i 为决策者 P_i 的支付函数,$\gamma_i = (\gamma_{i1}, \gamma_{i2}, \cdots, \gamma_{iN})$,$\gamma_{ij}$为决策者 P_i 在对策第 j 步时的策略. 在动态决策过程中,决策者所具有的信息结构的不同将导致不同的纳什解.

动态斯塔克伯格对策(dynamic Stackelberg game) 亦称主从对策. 有主从两方的对策. 属于非合作决策问题. 这一概念最初是由冯·斯塔克伯格(Von Stackelberg,H.)于 1934 年研究静态对策问题时提出的,故以斯塔克伯格命名. 这种对策的概念表述为:主方 P_1 选决策 $u_1\in U_1$,从方 P_2 选决策 $u_2\in U_2$,各自的支付函数为 $J_1(u_1,u_2)$和 $J_2(u_1,u_2)$. 对于 P_1 所选定的每一 u_1,P_2 均随之选 $u_2=T(u_1)$,以使对一切 u_2 有

$$J_2(u_1, T(u_1)) \leqslant J_2(u_1, u_2),$$

这里 T 是一个映射:$U_1\to U_2$. 然后,主方 P_1 再选 u_1^*,使对于一切 u_1 有

$$J_1(u_1^*, T(u_1^*)) \leqslant J_1(u_1, u_2).$$

于是这个 u_1^* 就是 P_1 的斯塔克伯格策略,而 $u_2^* = T(u_1^*)$就是 P_2 的斯塔克伯格策略. 在上述表达式中,以策略 $\gamma_i\in T_i$ 代替决策 $u_i\in U_i$,即得动态斯塔克伯格策略的定义.

鼓励控制策略(incentive control strategy) 一类动态斯塔克伯格对策问题中的一个概念. 它指主方根据自己的需要,选择并宣布一个奖惩函数以诱导从方采取某种有利于主方的策略. 这类问题的数学描述如下:记 P_0 为主方,$P_i(i=1,2,\cdots,M)$为从方,则问题为:求解 $\gamma_0: U_1\times U_2\times\cdots\times U_M\to U_0$,$\gamma_i\in T_0$,使得 $\arg\min\limits_{u_i} J_i(\gamma_0(u_1^t,\cdots,u_{i-1}^t,u_i,u_{i+1}^t,\cdots,u_M^t),$

$$u_1^t,\cdots,u_{i-1}^t,u_i,u_{i+1}^t,\cdots,u_M^t) = u_i^t (i=1,2,\cdots,M),$$
$$\gamma_0(u_i^t,\cdots,u_M^t) = u_0^t,$$

其中 $u_i^t(i=0,1,\cdots,M)$为主方所期望各方采取的决策. 满足上式的 γ_0 称为鼓励控制策略. 鼓励控制策略的意义在于:若从方不采取主方所期望的决策,则从方的支付将要增大,即受到惩罚.

动态协调问题(dynamic coordinate problem) 亦称协调问题. 一种微分对策. 各方协调以达到某种平衡的对策. 在合作对策中,是指在重复决策的每一过程中,决策者提出自己的策略,然后看对策结局是否对每一个决策者都满意. 若有不满意者,则再重复这一过程,直到所有决策者都满意为止,以使结局对大家都有利. 在动态非合作对策中,即是所谓"讨价还价"问题,是指在动态对策中,决策者通过谈判或协商来达到某种平衡. 宣布某一策略以阻止对方破坏对策过程中的平衡状态或一致状态,以使对策结

果对自己有利. 在对策之始，每一决策者都确定一个协调政策，一旦他们不能达到某种平衡，或达到了平衡以后又被对方破坏了，便采取这个协调策略的威胁部分. 用数学公式描述便是

$$\gamma_{ik}(x_0,\cdots,x_k)=\begin{cases}\bar{u}_{ik}, 若\ x_1=\bar{x}_1, x_2=\bar{x}_2,\cdots,x_k=\bar{x}_k,\\ d_{ik}(x_k), \quad 否则,\end{cases}$$

其中 $\bar{u}_{1k},\bar{u}_{2k}$ 为第 k 步时的平衡解，$\{x_0,\bar{x}_1,\cdots,\bar{x}_k\}$ 为相应的对策平衡状态，$d_{ik}(x_k)$ 是决策者 P_i 在第 k 步的威胁决策.

协调问题(coordinate problem)　即“动态协调问题”.

开环策略(open-loop strategy)　微分对策的术语. 指决策者具有开环信息结构的策略. 在 N 人动态对策过程中，如果决策者具有开环信息结构，则其所采取的策略称为开环策略. 开环信息结构之意为：在对策的每一步，决策者只了解对策初始时刻的状态，而不具有已经发生的每一步的信息. 开环策略一般有如下形式：

$$u_{ik}=\gamma_{ik}(x_0)\quad(i=1,2,\cdots,N),$$

其中 $k=1,2,\cdots,K$ 为决策步骤.

闭环策略(close-loop strategy)　决策者具有闭环信息结构的策略. 在 N 人动态对策过程中，如果决策者具有闭环信息结构，则其所采取的策略称为闭环策略. 闭环信息结构之意为：在对策的每一步，决策者对于此步以前的每一步对策状态的信息都了解. 此类策略一般有如下形式：

$$u_{ik}=\gamma_{ik}(x_0,x_1,\cdots,x_k)\quad(i=1,2,\cdots,N),$$

其中 $k=1,2,\cdots,K$ 为决策步骤.

还有一种闭环不完全信息结构，即在对策过程的每一步，决策者不能直接了解到以前每一步的状态，而是通过某种观测间接了解信息. 此类策略一般有如下形式：

$$u_{ik}=\gamma_{ik}(y_{i1},y_{i2},\cdots,y_{ik})\quad(i=1,2,\cdots,N),$$

其中 $k=1,2,\cdots,K$ 为决策步骤，

$$y_{ik}=h_{ik}(x_k),\quad k=1,2,\cdots,K$$

为决策者 P_i 在第 k 步时的观测.

反馈策略(feedback strategy)　微分对策术语. 指决策者具有反馈信息结构的策略. 在 N 人动态对策过程中，如果决策者具有反馈信息结构，则其所采取的策略称为反馈策略. 反馈信息结构之意为：在对策的每一步，决策者具有前一步对策状态的信息，即每一步根据前一步的状态来选择该步的策略. 一般有如下形式：

$$u_{ik}=\gamma_{ik}(x_k)\quad(i=1,2,\cdots,N),$$

其中 $k=1,2,\cdots,K$ 为决策步骤.

还有一种反馈不完全信息结构，即在对策过程的每一步，决策者不能直接了解到前一步的状态，而是通过某种观测间接了解信息. 此类策略一般有如下形式：

$$u_{ik}=\gamma_{ik}(y_{ik})\quad(i=1,2,\cdots,N),$$

其中 $k=1,2,\cdots,K$ 为决策步骤，$y_{ik}=h_{ik}(x_k)$ 为决策者 P_i 在第 k 步时的观测.

离散事件动态系统

离散事件动态系统(discrete event dynamic system)　简称 DEDS. 状态演化由发生于异步瞬时的离散事件触发的一类动态系统. 对 DEDS 的最早研究可追溯到对排队现象和排队网络的研究. 现今的研究兴起于 1980 年前后，由哈佛大学何毓琦教授提出，自此发展形成统一学科. 排队论、网络分析、计划评审、路径规划、调度排序等的研究对象都可归入 DEDS 范畴. DEDS 的工程背景，包括柔性生产线或装配线、计算机和通信网络、空中或机场交通管理系统、军事指挥自动化系统等高新技术问题. DEDS 本质上属人造系统范畴，其运行机制及研究方法都和连续变量动态系统有重要区别. 对 DEDS 的研究分为不同层次，分别采用不同系统模型和分析工具.

逻辑层次上，研究 DEDS 在逻辑时间下事件与状态符号序列关系的分析和控制，基本工具包括形式语言/有限自动机和 Petri 网等. 在自动机理论框架下，DEDS 采用有限自动机作为模型，其定义为一个五元组：

$$G=(Q,\Sigma,\delta,q_0,Q_m),$$

其中 Q 为有限状态集合，Σ 为有限事件集合，$\delta:\Sigma\times Q\to Q$ 为状态转移函数，$q_0\in Q$ 为初始状态，$Q_m\subset Q$ 为标识状态集合. 系统的演化是从初态 q_0 出发，由一个离散状态跳到另一离散状态，直到进入标识状态集 Q_m 而终止过程. 状态的跳变由发生的事件 $\sigma\in\Sigma$ 所触发，由转移函数决定跳变后状态 $q'=\delta(q,\sigma)$. 演化中的事件序列构成一个事件串，在 DEDS 分析和控制中扮演重要角色. DEDS 的分析归结为研究其逻辑行为，它由 G 的事件串集合表征，称为由 G 生成的子语言 $L(G)$. 事件集 Σ 中，有一些属于能控事件，即人为地可阻止其发生的事件，其集合为系统的可控事件集 Σ_c；另一些为不能控事件，其集合为系统的不可控事件集 Σ_u. DEDS 的控制称为监控问题，即综合监控器(也为一个自动机)，通过阻止某些能控事件的发生来排除不期望行为的出现，以使系统具有满意或最优性能. Petri 网同样可被用来对 DEDS 建模、分析和控制，并可通过引入时间、随机变量等因素来扩展模型的描述能力，因而愈来愈受到重视.

代数层次上，研究 DEDS 在物理时间下运行过程的分析和控制，基本工具包括极大、极小代数(或更一般的双子)等. 对一类确定性的 DEDS，通过引

入极大代数：

$\{\overline{R}=R\cup\{\infty\},\oplus=\max,\otimes=+,$零元$=-\infty\}$，

可将其状态 $x(k)$、输入 $u(k)$、输出 $y(k)$间的关系表征为一个“线性”模型：

$$x(k)=A\otimes x(k)\oplus B\otimes u(k),$$
$$y(k)=C\otimes x(k)\quad(k=1,2,3,\cdots).$$

在取状态反馈下，则得到闭环“线性”递推模型：

$$y(k)=M\otimes y(k-1)\quad(k=1,2,3\cdots).$$

平行于传统线性系统理论，在极大代数框架下，基于此“线性”模型可分析 DEDS 的特征结构（特征值和特征向量）、稳定性、周期性稳态、能达性、特征值配置、参量辨识等. DEDS 的控制，主要采取控制“事件发生时间”的途径，或更一般地基于事件图来综合.

统计性能层次上，研究 DEDS 在“随机情况”下的平均性能及优化，基本工具包括排队论、广义半马尔可夫过程等. 排队论框架下，DEDS 被抽象为“顾客等待和接受服务”的一个排队网络，它由若干服务中心按一定结构网络组成. 服务中心由顾客和一个或多个服务台构成，按约定顺序规则依次对到来顾客提供服务. 顾客分别按一定统计规律，进入某个服务中心等待并接受服务，完成服务后转移到另一服务中心，直至完成全部服务后离开网络. 研究的问题涉及不同服务规则、不同统计规律、服务台故障和修复下系统性能（输出率、利用率、等待时间等）的分析与优化，但只对极为简单情形才可得到解析形式分析结果. 摄动分析（PA）方法采用广义半马尔可夫过程描述 DEDS，由排队论方法结合计算机仿真来形成，首先通过对 DEDS 的一次仿真获取有关动态响应的标称样本轨道，进而由标称样本轨道运用简单分析手段构造出参数摄动下的摄动样本轨道，最后根据两类样本轨道并基于排队论方法计算系统性能相对于摄动参数的灵敏度，分析系统性能的优化.

DEDS 研究中一个突出方面是大多面临计算复杂性问题，为此已研发了一些有效的近似、次优和启发式算法，此外计算机仿真也是一个重要的手段. DEDS 的进一步发展，涉及混合动态系统，即由离散事件动态系统和连续变量动态系统的交互和耦合所构成的一类复杂系统，被认为是 DEDS 的一个新的生长点.

离散事件动态系统的监控理论（supervisory control theory of discrete event dynamic system） 简称 DEDS 监控理论. 逻辑层次上以自动机与形式语言为基本工具研究离散事件动态系统（DEDS）建模与控制的理论. 由拉马奇（Ramadge，P.）和旺纳姆（Wonham，M.）创立，常称为 R-W 理论. 在 DEDS 监控理论中，采用自动机描述受控对象和监控器，采用自动机的语言描述系统的行为.

自动机定义为一个五元组：

$$G=(Q,\Sigma,\delta,q_0,Q_m),$$

其中 Q 为有限状态集，Σ 为字母表即有限事件集，$\delta:\Sigma\times Q\to Q$ 为状态转移函数，$q_0\in Q$ 为初始状态，$Q_m\subset Q$ 为标识状态集. 自动机 G 的语言定义为字符串集合

$$L(G)=\{\sigma^*\in\Sigma^*\mid\delta^*(q_0,\sigma^*)\in Q\},$$

自动机的标识语言定义为字符串集合

$$L_m(G)=\{\sigma^*\in\Sigma^*\mid\delta^*(q_0,\sigma^*)\in Q_m\},$$

其中 Σ^* 为中有限字长字符串（包括空字符串）的集合，$\delta^*:\Sigma^*\times Q\to Q$ 和 $\delta^*:\Sigma\times Q\to Q_m$ 为状态转移函数，分别可通过递归由 δ 来定义.

对受控自动机 G，Σ 可分解为可控事件集 Σ_c 和不可控事件集 Σ_u，$\Sigma=\Sigma_c\oplus\Sigma_u$，$\Sigma_c\cap\Sigma_u=\Phi$（空集），定义 $\Gamma=\{\gamma\mid\gamma:\Sigma\to\{0,1\},\gamma(\sigma_u)=1,\sigma_u\in\Sigma_u\}$为控制模式集. 基本监控方式有事件反馈监控器和状态反馈监控器两类. 事件反馈监控器由监控自动机 $S=(X,\Sigma,\xi,x_0,X_m)$及其状态集到控制模式集的映射 $\varphi:x\to\Gamma$ 所构成，闭环系统为

$$S/G=\{X\times Q,(\xi\times\delta)^f,(x_0,q_0),X_m\times Q_m\},$$

其中，当映射 $\varphi(x)\sigma=1$ 时，$(\xi\times\delta)^f(\sigma,x,q)=(\xi(\sigma,x),\delta(\sigma,q))$；当映射 $\varphi(x)\sigma=0$ 时，$(\xi\times\delta)^f(\sigma,x,q)$ 无定义. 状态反馈监控器为 $f:Q\to\Gamma$，闭环系统为

$$G^f=(\Sigma,Q,\delta_c^f,q_0,Q_m),$$

其中，当控制模式 $\gamma(\sigma)=f(q)(\sigma)=1$ 时，$\delta_c^f(\sigma,q)=\delta(\sigma,q)$；当控制模式 $\gamma(\sigma)=f(q)(\sigma)=0$ 时，$\delta_c^f(\sigma,q)$ 无定义. DEDS 监控问题是：对给定一个受控自动机 G 和目标语言 $L\subset L_m(G)$，综合一个监控器（事件反馈(S,φ)或状态反馈 f），使闭环系统标识语言属于目标语言（$L_m(S/G)\subset L$ 或 $L_m(G^f)\subset L$），且尽可能为最大.

在 DEDS 监控理论中，研究的问题包括能控性和监控器的存在性、最大能控语言、能观测性和观测约束下的监控器、分散监控和分散化监控、允许阻塞监控和赋时自动机监控等. 监控理论在处理标准逻辑规范范围内监控问题上已形成系统和完整的理论，但在处理实时、并发等更复杂规范范围内监控问题上则有待于系统和成熟.

Petri 网（Petri Net） 用以研究异步并发系统的组织结构和动态特性的一类图形化方法. 简称 PN. Petri 网的应用领域遍及计算机软硬件、分布式数据库、生产管理、通信网络、柔性制造系统等的建模和分析. Petri 网是在逻辑时序下研究离散事件动态系统（DEDS）特性和性能的一种有效的手段.

Petri 网定义为一个六元组

$$\mathrm{PN}=(P,T,F,W,M,M_0),$$

其中 $P=\{P_1,P_2,\cdots,P_m\}$为有限位置集，$T=\{t_1,t_2,\cdots,t_n\}$为有限变迁集，$F\subseteq(P\times T)\cup(T\times P)$为节点

流关系集即有向弧集,$W:F\to\{1,2,\cdots\}$为有向弧的权,$M:P\to\{0,1,2,\cdots\}$为状态标识,$M_0:P\to\{0,1,2,\cdots\}$为初始标识. 在 Petri 网的图形表示标识 Petri 网图中,$\{p_1,p_2,\cdots,p_m\}$取为位置节点,用"◦"(圈)表示;$\{t_1,t_2,\cdots,t_n\}$取为变迁节点,用"—"(粗线段)表示;F 构成有向弧集,从位置节点 p 指向变迁节点 t 的有向弧表示为(p,t),从变迁节点 t 指向位置节点 p 的有向弧表示为(t,p);W 的元构成相应有向弧的权;$M(p_i)$为位置节点 p_i 中"托肯"数(黑点数),行向量 $M=[M(p_1),M(p_2),\cdots,M(p_m)]$代表状态标识;$M_0$ 对应于状态标识 M 的初始设置. Petri 网本身仅能描述系统的静态结构,动态过程需要由状态标识的变化迁移过程来表征. 变迁发射规则是 Petri 网中最基本的约定,包括变迁发射的和发射后完成的操作的规则,可用来决定状态标识能否变迁和变迁后系统的状态标识.

对 Petri 网的分析主要围绕可达性、活性、死锁、有界性、安全性等基本特性. 分析方法主要有可达树法和矩阵方程法,分别从图和代数角度建立了特性与结构间的关系. 现今,在实际需要的推动下,已经对 Petri 网的功能和属性从不同侧面作了拓展. 为研究控制问题,在 Petri 网中引入可控位置集和相应输入弧集以构成受控 Petri 网——CPN,综合方法有谓词方法和代数方法. 为克服对大规模 DEDS 建模中维数过大的问题,引入具有集结特性的着色 Petri 网. 为反映变迁发射中的时延和时延的随机性,分别引入赋时 Petri 网和随机 Petri 网. 为解决大规模 Petri 网特性分析中的计算量问题,引入基于子网分析的网分解方法. 为扩展同时包含离散事件和连续变量的系统,引入混合 Petri 网. Petri 网方法(特别是可达性和活性分析)中面临的主要问题是计算复杂性,不少问题属于 NP 完全问题.

Petri 网是由德国学者派垂(Petri,C. A.)于 1962 年提出的.

离散事件动态系统的代数方法(algebraic method for discrete event dynamic system)　简称 DEDS 的代数方法. 研究离散事件动态系统的有关理论. 指动态时间性能层次上以极大代数或/和极小代数为基本工具研究离散事件动态系统(DEDS)建模、分析与控制的理论. 极大代数是一种双子代数,其定义域为 $\overline{R}=\mathrm{R}\cup\{-\infty\}$,R 为实数域;两种基本运算为加法$\oplus$和乘法$\otimes$:$a,b\in\mathrm{R}$,$a\oplus b=\max\{a,b\}$,$a\otimes b=a+b$;$\varepsilon=-\infty$和 $e=0$ 分别为零元和单位元. 对应地,若取定义域为 $\overline{R}=\mathrm{R}\cap\{+\infty\}$,R 为实数域,基本运算加法$\oplus$和乘法$\otimes$定义为:$a,b\in\overline{R}$,$a\oplus b=\min\{a,b\}$,$a\otimes b=a+b$,零元和单位元分别为 $\varepsilon=+\infty$和 $e=0$,则构成极小代数.

基于极大代数,可把制造系统中的串行生产线等一类广泛 DEDS 的原始逻辑非线性状态方程和输出方程实现线性化,表示为线性 DEDS:

$$x(k)=A\otimes x(k)\oplus B\otimes u(k),$$
$$y(k)=C\otimes x(k)\quad(k=0,1,2,\cdots).$$

平行于域上的线性连续变量系统,可由此来研究线性 DEDS 的分析和综合. 系统的分析主要集中在特征结构(特征值和特征向量)、稳定性、周期性稳态、能控性和能观测性、事件序列性、传递函数、最小实现、鲁棒性等. 系统的综合包括状态反馈、极点配置、周期参量配置、控制综合等. 线性 DEDS 的极大代数方法可区分为基于状态方程的时间域方法和基于传递函数的频率域方法.

基于研究复杂类型逻辑电路等复杂 DEDS 的需要,通过同时引入"极大 $\vee=\max$,极小 $\wedge=\min$,乘法$\otimes$"运算,形成 DEDS 代数方法中一个新的模型,即双子代数上的一类非线性 DEDS,研究问题包括稳定性、周期性稳态、鲁棒周期性等. 非线性 DEDS 在研究难度上要较线性 DEDS 为大.

摄动分析(perturbation analysis)　简称 PA. 关于系统性能优化的方法之一. 指基于排队网络分析和计算机仿真相结合研究随机离散事件动态系统(DEDS)性能优化的一种简化实用方法. 由哈佛大学何毓琦教授在 20 世纪 80 年代初提出. 随机 DEDS 广泛存在于制造系统、计算机网络、通信系统、交通系统等中,其模型为排队网络. 随机性能优化问题具有很高的复杂性,需要很大的计算量.

PA 采用广义半马尔可夫过程(GSMP)作为系统模型. PA 的思路可归纳为:首先,通过对标称参数系统的一次计算机仿真或实验,观测记录有关动态响应的一个样本实现,来得到标称样本轨道,以及其形式化描述即系统状态方程模型. 进而,在选定参数的变动下,利用相似性条件的简单摄动传播规则,基于计算分析的途径,由标称样本轨道构造出摄动样本轨道. 最后,基于系统标称样本轨道和摄动样本轨道的结果,计算出系统性能相对于变动参数的灵敏度即性能优化估计. 与随机 DEDS 的仿真方法相比,PA 只需一次仿真,使计算机仿真机时大为减少;与排队网络分析方法相比,PA 只需要很弱的假设就能够处理非常一般的排队网络问题,并可使计算量得到显著减少.

设可变动参数为 θ,参数变动为 $\Delta\theta$,对应于样本 ω,标称轨道的标称性能为 $PM(\theta,\omega)$,参数变动后的摄动性能为 $PM(\theta+\Delta\theta,\omega)$,PA 方法可给出一次样本轨道的性能 $PM(\theta,\omega)$的导数值 $\partial PM(\theta,\omega)/\partial\theta$. 但对随机过程,更有意义的分别是样本轨道性能 $PM(\theta,\omega)$ 及其导数值 $\partial PM(\theta,\omega)/\partial\theta$ 的均值 $E[PM(\theta,\omega)]$和 $E[\partial PM(\theta,\omega)]/\partial\theta$. 对于样本性能

$PM(\theta,\omega)$可得到解析表达式的情况，只有当$PM(\theta,\omega)$对参数θ的灵敏度估计满足无偏性，也即

$$E[\partial PM(\theta,\omega)/\partial\theta] = \partial E[PM(\theta,\omega)]/\partial\theta$$

（“均值算子”和“偏导数算子”可变换运算顺序）成立，才可由样本性能导数值$\partial PM(\theta,\omega)/\partial\theta$来估计均值性能导数值$\partial E[PM(\theta,\omega)]/\partial\theta$. 对于样本性能$PM(\theta,\omega)$不能得到解析表达式的情况，只有当$PM(\theta,\omega)$对参数$\theta$的灵敏度估计满足强一致性，才可由样本性能导数值$\partial PM(\theta,\omega)/\partial\theta$来估计均值性能导数$\partial E[PM(\theta,\omega)]/\partial\theta$.

针对基本PA方法要求满足相似性条件和满足无偏性或强一致性的问题，已经相继发展了推广摄动分析和平滑摄动分析. 推广摄动分析适用于不满足相似性条件的排队问题. 平滑摄动分析可对一些简单的DEDS使样本性能灵敏度估计做到无偏或强一致.

离散事件动态系统仿真（simulation for discrete event dynamic system） 简称DEDS仿真. 基于模型通过软件或程序在计算机上按数值方法求解离散事件动态系统(DEDS)动态过程问题的一种方法. 早在20世纪的70—80年代，就已经出现商用的DEDS仿真软件，如SIMAN, SIMSCIPT, SIMULA, GASP, GPSS等，可应用于制造系统、交通运输系统、服务系统等的调度，但难以实现复杂DEDS的随机优化仿真. Petri网作为具有很强描述功能的系统模型，也常被用于计算机系统、制造系统等复杂类型DEDS的仿真中.

DEDS仿真有着广泛的应用. DEDS动态性能优化问题的摄动分析(PA)方法和次序优化(ordinal optimization)方法，就是建立在随机DEDS仿真的基础上的. 在随机DEDS仿真中，常采用广义半马尔可夫过程(GSMP)作为系统模型. 考虑一个随机过程$x(t), t\in[0,+\infty)$，设其状态空间X和事件集E均为有限或可数集，$E(x)\subset E, x\in X$为对应于状态x的有效事件集，且对每个$e\in E$其生命周期$l(e)$是分布函数为Φ_e的随机变量. 若对时刻t和状态x，每个$e\in E(x)$的剩余生命时间为$l_r(e)$，有下一个状态转移发生的时刻为

$$t^* = \min_{e\in E(x)} \{t + l_r(e)\},$$

触发事件为e^*，触发时刻为$t^* = t + l_r(e^*)$，并以转移概率$P(x';x,e^*)$转移到状态x'，则称$x(t)$为基于(X,E,P,Φ_e)的一个GSMP.

对广义半马尔可夫过程$x(t)$的仿真，在摄动分析中即为获取标称样本轨道，其步骤为：

1. 初始化，以概率P_0选择初始状态$x_0 = x(t_0)$.
2. 令$n=1$.
3. 令$t = n\Delta t$.
4. 计算$x(t)$和$E(x)$.
5. 计算$l_r(e), \forall e\in E(x)$.
6. 计算下一状态转移时刻$t^* = t + l_r(e^*)$、触发事件e^*、触发时刻$t^* = t + l_r(e^*)$.
7. 以转移概率$P(x';x,e^*)$随机采样，得到转移后新状态x'. 常用的采样方法为逆函数法，设随机变量为ξ，其分布函数$F(\xi)$的逆存在，则$\xi = F^{-1}(u)$，u为$[0,1]$上均匀分布的随机变量，由此通过对u的采样可得到ξ的采样.
8. 令$n = n+1$.
9. m为设定的仿真周期. 若$n\leqslant m$，转移到3；若$n>m$，仿真结束.

对于摄动分析，得到的是DEDS的标称样本轨道，可用来构造摄动轨道以及进一步计算性能优化.

计算机辅助设计

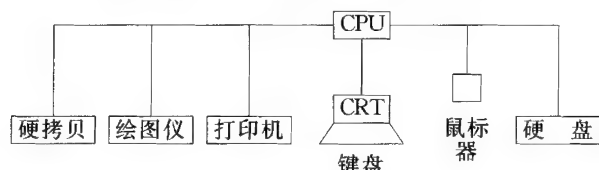
控制系统计算机辅助设计（computer aided design in control system） 简称CADCS. 利用计算机作为辅助工具，进行控制系统设计的技术. 计算机辅助设计都是有对象的. 例如，用计算机设计机器零件、模具等，称为机械计算机辅助设计；用计算机设计房屋，称为建筑计算机辅助设计；同样，用计算机设计控制系统，称为控制系统计算机辅助设计. 当控制理论研究有了一些结果，可以用CADCS软件进行仿真，使其向实际应用跨进一步，或者针对实际的控制系统，选择适当的控制理论方法进行仿真，进而解决实际问题. 同时，对仿真过程与结果的分析，往往又进一步提出控制理论问题，促进理论的发展. CADCS是控制理论与实际应用相结合的不可缺少的现代技术. 20世纪50年代末期以来，由于航天技术的推动，状态空间方法得以发展，引入了复杂的数值计算，大量的数据处理与决策过程，迫切需要使用计算机，于是60年代在控制系统设计上首先用到了软件的子程序. 在这种情况下，刚刚跨入70年代就出现了CADCS软件包，并且在十年之内有了一个大发展. 典型代表有两个：一个是英国曼彻斯特大学罗森布罗克学派的现代频域法软件包；一个是瑞典隆德(Lund)大学奥斯特隆姆(Astrom, K. J.)和他的学生们的状态空间方法的一组软件包：IDPAC, INTRAC, MODPAC, POLPAC, SIMNON, SYNPAC，使其在国际上处于很有影响的地位. 80年代初期，美国莫勒(Moler, C.)开发了MATLAB，它的数据结构是复矩阵，MATLAB就是完成各种矩阵运算，其特点是有一个数学化的人机交互界面. 这是一个很好的线性系统计算机辅助设计开发工具，在国际上有很大影响. 80年代末到90年代初，CADCS已从直接开发软件包转入研究CADCS开发工

具与开发环境，目前，已经有一些可喜的研究成果.

在中国，CADCS 的研究始于 20 世纪 70 年代末 80 年代初，在 80 年代得到很好的发展，典型代表是 CADCSC. 目前，中国学者也正在从事 CADCS 开发环境的研究工作.

对于 CADCS 的研究工作，一方面必须对控制理论及其算法有深入的了解，另一方面必须掌握现代软件设计方法，这是两大基础，其创造性的工作在于两者的高度结合.

CADCS **硬件**(CADCS hardware)　支持 CADCS 工作的计算机及其有关的外部设备. CADCS 系统的硬件如下图所示，可以由个人用微型计算机或小型计算机配置适当的输入输出设备构成. 也可以采用 CAD 工作站，例如486 SUN 工作站等，必须配备适当的软盘及大容量的硬盘. 另外，视需要配备彩色 CRT、鼠标器、打印机、绘图仪、硬拷贝等.



CADCS 的硬件示意图

CADCS **支撑软件**(CADCS support software)　支持 CADCS 设计的系统软件. 在确认的机型下，选择适当的操作系统与算法语言. 例如，在 486 下，可选择 DOS、Windows 操作系统，以及 C 语言、VisualC++ 语言. 另外，非线性函数表达式的求值、求导软件可直接支持 CADCS 的总体系统设计；矩阵运算、优化算法、多项式阵算法、统计算法软件决定了 CADCS 的数值质量；甚至，还有更接近 CADCS 设计功能的软件. 例如模型转换、模型简化、系统仿真等软件，它们本身已具有 CADCS 的设计功能，同时又为 CADCS 更广泛的设计功能提供了一个支撑软件. 随着计算机软件的发展，CADCS 的支撑软件会越来越好.

控制理论算法(control theory algorithm)　控制理论的数值计算或推理方法. CADCS 是一个软件，它所表达的内容是控制理论提出的问题、方法与结果. 控制理论的任何方法被 CADCS 表达时，必须给出算法，这就是控制理论算法. 当前，这种算法大量地表现为数值算法，但不一定非得数值算法，只要从问题出发，经过一定的有限程序步骤得到明确结果的过程都是算法. 解决控制理论的同一个问题，可以有很多不同的算法，这方面的论文、著述很多，可以说琳琅满目. 控制理论工作者的任务之一是寻求好的算法. 一个算法在什么条件下是好的，这是一个极富挑战性的问题，也是 CADCS 设计者希望得到解决的问题.

CADCS **基础算法**(CADCS basic algorithm)　被控制理论算法引用的通用性计算方法. 控制理论算法中，常常用到一些已知的算法，这些算法有独立意义，常常被不同的控制理论算法所引用，其本身绝大部分在 CADCS 中不具有控制理论算法的意义，这些算法被称之为 CADCS 基础算法. 例如，矩阵运算、优化算法、多项式阵运算、统计算法等. 这些算法的好坏直接影响控制理论算法的 CADCS 中的时间复杂性与空间复杂性，以及精度等数值质量. 这些基础算法的研究都是计算数学的主要研究方向，计算数学可为控制理论提供好的基础算法. 反过来，控制理论算法的研究也可给计算数学提出富有启发性的基础算法问题. 类似控制理论算法，一个基础算法在什么条件下是好的，同样也是一个极富挑战性的问题.

CADCS **语言**(CADCS language)　表达 CADCS 的专用语言. CADCS 语言是进行控制系统计算机辅助设计的专用语言，它在 CADCS 领域内通用，除了控制运算顺序与通用算法语言的类似外，其他内容均有独特的要求，相应于数据类型的是控制系统的数据结构，相应于表达式的是控制系统建模、辨识、预报、分析、综合与仿真的算法，相应于输入输出的是 CADCS 人机交互界面. 从 20 世纪 70 年代中期算起 20 多年来的 CADCS 软件中，已给出若干种命令式语言，例如 IDPAC，MATLAB，CADCSC 等，它们都表达了当时的水平，现在还在发展. 早在 1984 年奥斯特隆姆(Astrom，K. J.)对 CADCS 命令语言就提出过明确要求：完整性，即覆盖大范围的 CADCS 算法；较强的表达力，用户可用命令描述自己想作的工作；少而精，易学也易记忆，权衡与强表达力的关系；反映理论的自然含义，使控制理论工作者易于使用. 进一步发展这种语言的设计，对 CADCS 的研究者们是一种具有强大吸引力且困难的工作.

CADCS **人机交互界面**(CADCS interface)　专为 CADCS 用的人机交互界面. CADCS 是一个人机系统，人向机器提出问题，选择或输入参数，从机器得到设计结果，这样的人机交互界面，要求内容正确、形式便于操作、阅读和应用. 早期的人机交互界面是人机对话式的，常见的人机交互界面为选单式与命令式. 选单式人机交互界面易于研制，有详尽的提示，对初学者或初次使用者方便，但重复进行辅助设计时，冗余信息过多；命令式人机交互界面进行重复设计时方便，但初次使用时，需记忆的内容过多. 混合式人机交互界面包含选单式和命令式人机交互界面，在选单式人机交互界面下工作，可以自动生成命令，在命令式人机交互界面下工作，可以回显选单

的提示，作为对命令的注释. 例如 CADCSC 就具有混合式人机交互界面，它保留了选单式和命令式人机交互界面的优点，同时克服了它们的缺点.

CADCS **系统结构**(CADCS system structure) 用来描述控制系统的数据结构. 在 CADCS 中，对一个控制系统，总可以找到一种方法对它进行描述. 不管对线性系统还是非线性系统，方法之一就是将系统分为系统结构和系统参数. 对于线性系统，系统结构可以用系统结构参数描述，例如用差分方程表示的多输入-单输出系统，其系统结构参数可为输出阶数、输入个数、各输入的阶数和滞后、噪声的阶数和强度. 当系统参数确定后，这个系统就会得到完全的描述. 对于非线性系统，系统结构可以用一组非线性函数表达式表示，例如 $y=f(u,x,p)$，输出 y 是明确的，不管 f 与输入 u、状态 x 以及系统参数 p 的关系如何，其系统结构都可以用一个非线性函数表达式 f 描述. 当给出系统参数 p 时，这个系统也会得到完全的描述. 有的系统如果其系统结构既含有若干非线性函数表达式，也含有系统结构参数，可称为混合结构. 例如，最优控制问题的系统结构就可以用混合结构描述. 随着控制理论的发展，还会提出新的系统结构，CADCS 工作者的任务之一是不断探求系统结构描述的一般形式.

CADCS **可调量**(CADCS adjustable variable) 在人机交互过程中可修改的量. 在 CADCS 中，对系统建模、辨识、预报、分析、综合与仿真等辅助设计过程中，一般地，在每一个辅助设计算法中，都有一些在人机交互过程中可供选择或修改的量，称为可调量. 当可调量变化时，重复进行辅助设计，以期获得好的设计结果. 可调量的性能描述，应该使其在人机交互过程中易于使用. 方法之一是给出可调量的名称、类型、取值范围、缺省值与当前值. 在 CADCS 的人机交互过程中，可调量处在最内层的人机交互界面上，可在适当的窗口技术支持下，构成这个可调量人机交互界面.

CADCS **子系统**(CADCS subsystem) 表达 CADCS 总体系统的一部分功能. 当 CADCS 软件系统过大时，可按内容分类或按控制系统设计与使用方便分类，称每一类为一个子系统，例如 CADCSC 软件系统就是按内容分成若干个子系统. 由若干个子系统集成为一个 CADCS 软件系统时，要考虑软件系统对子系统的驱动与控制、系统与子系统之间以及不同子系统之间的通信、功能命令与人机交互界面的一致性等. 对已分别独立设计的子系统进行集成是一件很困难的事情，对功能与人机交互界面很难给出好的结果. 如果进行总体系统设计时，就考虑到子系统的结构及其可扩展性，那就会有一个好的集成效果. 但是，由于其总体系统已被确定，当扩展子系统时，还会受到局限，比较好的方法是设计一个自动生成系统，可按实际情况生成满足要求的子系统.

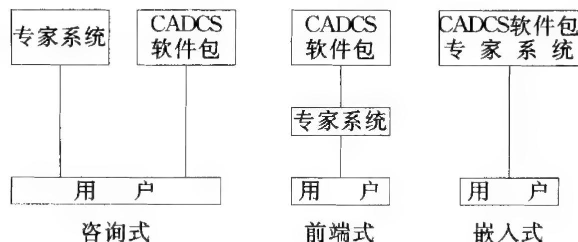
CADCS **数据库**(CADCS database) 表达控制系统及其运算与通信的数据库. 数据库是从生成报表开始的，描述了报表的数据结构及其上的相应运算，例如，dBASE，FOXBASE，FOXPRO，SYBASE，ORACLE 等，对财务、仓库、人事、计划、技术、生产、供销等管理有效. 也曾有人将其引入 CADCS 中使用，当然不易成功. 确切地说，这个数据库应该称为报表数据库. CADCS 数据库描述的对象是控制系统，并在给定的控制系统上引入相应的运算，这些运算包括控制系统设计、内部通信与基本外部通信. 可应用 CADCS 数据库构成 CADCS 开发环境的工具之一.

控制理论化的人机交互界面(interface of control theory) 用图形直接描述控制系统原始表达形式的人机交互界面. 到目前为止，CADCS 的人机交互界面都是在一般的计算机软件人机交互界面基础上发展起来的，将控制系统及其设计结果的描述分解成不同类型的数据、表达式与图形. 在窗口与图形技术的支持下，构成的人机交互界面，虽然获得很大的成绩，但并不理想. 在软件领域往往希望得到“自然语言”表述的人机交互界面，在 CADCS 领域内也希望设计“图形化”的人机交互界面，前者是难以实现的理想状态，后者是可以实现的，但仅提到形式，没涉及内容. 奥斯特隆姆(Astrom，K. J.)在描述 CADCS 语言时，曾提出语言应反映理论的自然含义，对 CADCS 人机交互界面来说，无疑也是正确的. 显然，应该考虑面向对象的人机交互界面，这个界面可用图形的形式直接描述控制系统的原始表达形式，这也正是在控制理论范围内的局部自然语言人机交互界面，称其为控制理论化的人机交互界面.

矩阵实验室(matrix laboratory) 一种著名的矩阵计算软件包. 1980 年，美国莫勒(Moler. C)开发了 MATLAB，由 Mathwork 公司推出，它的数据结构是复矩阵. MATLAB 就是在复矩阵上完成各种矩阵运算. 人机交互界面是数学表达式形式的命令语言. 例如 $C=A\times B$，表示矩阵 A 与矩阵 B 相乘，结果送入矩阵 C. 同时给出相应的矩阵输入与编辑系统，它不但有行的概念，而且还有列的概念. 国际上，多年来矩阵运算形成了两个著名的软件，就是 LINPACK 与 EISPACK. MATLAB 是在这两者的基础上发展起来的，它的贡献在于它有一个好的数学化的矩阵运算人机交互界面. MATLAB 的出现，在世界各地受到了极大的重视，凡是建立在矩阵运算基础上的科学与工程技术数值计算问题，都获得了极大的方便. 在 CADCS 线性系统领域也是如此.

CADCS 算法工具(CADCS algorithm tool) 一个比一般算法语言更易于表达CADCS算法的软件.对线性系统的CADCS,MATLAB就是一个很好的算法工具,它比任何高级算法语言都更易于表达线性系统的CADCS算法.对一般的CADCS,仅仅矩阵运算是不够的,必须引进新的算法,同时新算法层出不穷,比较好的方法是给出一个既能包容已有的算法,也可随时引入新算法的工具.另外,当用MATLAB设计一个CADCS软件时,不能给出好的CADCS人机交互界面,相应的数据结构也受到了限制.解决这些问题的方法之一是将已建立起来的算法工具嵌入一个高级算法语言中,例如,将MATLAB嵌入TurboC中形成的M-TurboC就是这样的CADCS算法工具之一.一般地,一个比已有的算法语言更易表达CADCS算法的软件称为CADCS算法工具.

CADCS 专家系统(CADCS expert system) CADCS范围通用的专家系统.如图.常见的CADCS专家系统有三种类型:咨询式,用户将控制系统设计问题输入到专家系统,专家系统运行后,告诉用户使用CADCS软件包的方法,用户应用CADCS软件包得到所需求的结果;前端式,用户将控制系统设计问题输入到专家系统,专家系统自动启动CADCS软件包,并将得到的结果告诉用户;嵌入式,专家系统被嵌入到CADCS软件包中,成为该软件包的有机组成部分,依用户输入的问题运行不同的命令序列,在命令运行时,选择不同可调量的组合及其恰当的值,以期得到设计结果.为了提高设计CADCS专家系统的效率,可设计一个CADCS通用专家系统,作为CADCS的工具之一.



常见的专家系统示意图

CADCS 开发环境(CADCS environment) 开发CADCS软件包的一套工具组合软件.CADCS开发环境是当前国际上CADCS研究工作的热门课题.方法之一是分别开发控制理论化的人机交互界面、CADCS数据库、CADCS算法工具、通用的CADCS专家系统,使其在统一模式下构成一个可对CADCS软件包进行组态的软件系统,这就是CADCS开发环境.运用这个开发环境,可依个人的意愿,很容易地构成自己的CADCS软件包.特别地,当控制理论工作者得到新的控制理论算法时,将会感到CADCS开发环境是一个好帮手.

系统辨识软件包(identification package) 一个系统辨识软件包.在20世纪70年代CADCS大发展的时期,典型代表之一是IDPAC.国际上自动控制领域中最知名的学者之一——瑞典隆德大学的奥斯特隆姆(Astrom,K.J.)与他当时的学生们花费了10年的心血,用状态空间方法发展了多变量系统控制理论.在软件包的使用与结构上克服了刻板地提问与回答的限制,提出了命令式人机交互界面,在控制系统设计中给系统设计者以主动权,将交互过程提高到一个新阶段,完成了由IDPAC,INTRAC,MODPAC,POLPAC,SIMNON与SYNPAC等六个包组成的一套软件包,使其在国际上处于很有影响的地位.IDPAC是一个系统辨识软件包,不仅体现了他们设计上的特点,在应用上也获得了很好的效果.这个软件包对中国CADCS的研究工作有重大影响.例如,辨识、控制与仿真软件包(ICS)、线性与非线性系统辨识软件包(LANSIP)与CADCSC软件系统的设计都是在IDPAC的基础上发展起来的.

在线的CADCS (CADCS on line) 可用于实际控制现场进行实时控制的软件包.一个CADCS软件包,如果可以从控制现场直接取得检测数据,并可将控制信号输出到现场,这就是一个在线的CADCS软件包.它可以在线地进行系统建模、辨识、预报、分析、综合、仿真,也可建立专家系统,必要时可进行现场控制实验.当前DCS(distributed control system)在工业控制中已占居主导地位,DCS的设备生产销售公司已推出成熟的常规组态软件,但现代控制理论进入现场控制还是一件很困难的事情,如果将CADCS引入DCS,使在线的CADCS得到发展,将CADCS与DCS一体化,那将会推动控制理论的应用,反过来,也会推动控制理论本身的发展.

中国控制系统计算机辅助设计(computer aided design in control system of China) 中国控制系统计算机辅助设计软件系统.这是一个由中国专家设计完成的CADCS软件系统.CADCSC可以解决单变量、多变量、连续、离散、线性、非线性控制系统的建模、辨识、预报、模型转换、分析、综合、仿真等问题,对系统的分析与设计包括状态空间方法、现代频域法与多项式阵方法,另外还有单变量、多变量与模型参考自适应控制、最优控制、奇异摄动线性系统、快速跟踪等.其中好多方法是中国学者独立提出与发展的.CADCSC有如下特色:内容丰富;具有表达力强且少而精的CADCSC语言;友好的混合式人机交互界面;统一的子系统过程与子系统特征常量;完

备的线性系统结构分类;一类非线性系统的表达与处理系统;统一的低层可调量处理系统;可靠的基础库与CAD算法的层次结构CADCSC为其非线性控制子系统填补了一项空白.

CADCS **的评价**(evaluation of CADCS) 用一些客观标准去度量一个CADCS软件好坏的方法.评价一个CADCS软件的好坏是一件很重要的工作,它涉及CADCS的发展方向.一般地,要从CADCS语言、人机交互界面、数学算法与CADCS算法、数据结构、智能特征、对教学的适应性、是否可在线运行等诸方面去评价.总的来说,是要看控制理论与计算机技术结合的程度与应用水平.结合的程度越高,应用越广,其水平就越高.

CADCS **的基准检测程序**(benchmark of CADCS) 检测用于控制系统计算机辅助设计的各类基本计算方法和所使用方法的正确性等的基准程序.控制系统计算机辅助设计属于技术科学的范畴,它的一个特点是对于同样的控制对象,可以有多种设计方法.即使对于同一种设计方法,还可以有多种计算方法.控制系统计算机辅助设计的基准检测程序分为三类.

第一类是检测用于控制系统计算机辅助设计的各类基本计算方法的基准程序.例如,对于求矩阵的特征值问题,有罗塞矩阵;对于多维优化问题,有罗森布罗克函数;对于常微分方程的初值问题,有范得波尔方程.它们的特点是其结果是知道的,但是当在计算机上解这些问题时,相应的条件很差.因而,如果一种计算方法能够在可接受的计算时间内以足够的精度求解这些问题,就可以认为它们能适应一般的控制系统计算机辅助设计问题.

第二类是检测控制系统计算机辅助设计所使用的方法的正确性.由于控制系统以及设计要求的多样性,一种设计方法通常只能用于某一类控制系统.目前,在提出一种设计方法时,需要同时提出相应的检测例子.例如,罗森布罗克(Rosenbrock,H.H.)在提出多输入-多输出系统的频率域设计方法时,给出了一个将此方法用于设计一个飞机发动机控制的例子.萨方诺夫(Safanov)在提出用求解两个黎卡提方程来进行H_∞鲁棒设计时,也给出了一个飞机控制与一个大型结构控制的检测例子.这些例子说明所提出的方法对于某一类控制问题是有效的,但不能表明此种方法对其他控制问题也是有效的.这乃是当前设计控制系统计算机辅助设计基准检测程序的主要困难.

第三类是提出一些典型的控制问题,用各种不同的设计方法进行设计,并对设计结果进行比较,依此对各种设计方法的适用范围进行评价.和第二类检测程序不同,对于参与检测的参加者来说,系统的精确模型是不知道的.例如,在1993年国际自动控制联盟世界大会的准备过程中,由格莱比(Graebe,S.F.)向有兴趣的研究者提供了一个称为IFAC-93的检测程序.它是仿真一个对象的标准的C源码.用户可在任一时刻给出一个控制信号,该码就仿真下一时刻的输出响应.此对象是时变的,检测程序提供了三个“强度水平”,它越高表明时变越强.此码是加密的,没有人能读懂它.检测程序不提供对象的阶和参数值,但允许通过仿真实验来尽可能多地学习.参加者可以自由地编制界面和插入任何固定的或适应的控制器,且由编译该码和仿真相应的结果来调整控制器.有国际上九个著名的研究机构参加了这次检测,他们分别用他们各自的研究成果来进行设计,并在1993年的IFAC世界大会上交流了设计结果.通过这次检测,对于目前在鲁棒控制和适应控制中提出的众多方法给出了适当的评价.例如:

1.对于小的对象变化,H_∞控制器比经典的比例-积分控制器只好一点点.此时适应控制器可能带来不必要的品质降低.

2.对于中等的对象变化,H_∞控制器比其他定常的控制器好,这是其复杂性带来的最大利益.此时适应控制器也有可能得到好结果,看来调整是重要的.

3.对于大的对象变化,没有定常的控制器能与好的适应控制器相比.

撰　稿　王　龙　王连文　王金枝　王炎生　王治宝　王恩平　王朝珠　井元伟　叶庆凯　冯昭枢　冯德兴　刘永清　安鸿志　安森建　许可康　李训经　李春文　吴　麒　吴智铭　何关钰　邹　云　张　霖　张纪峰　张嗣瀛　陈　辉　陈光亚　陈兆宽　陈宗基　陈树中　陈祖浩　陈彭年　陈翰馥　郁文生　郑　锋　郑大钟　胡启迪　俞文焕　洪奕光　秦化淑　袁著祉　耿志勇　夏小华　高为炳　高黛陵　郭　雷　席在荣　涂菶生　黄　琳　韩正之　程　勉　程代展　褚家晋　霍　伟

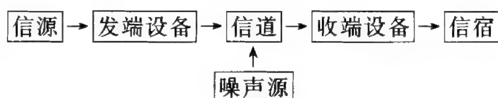
审　阅　王　龙　冯德兴　刘永清　吴　麒　陈宗基　郑大钟　秦化淑　黄　琳　霍　伟

通 信 与 信 息 理 论

通 信 总 论

通信(communication) 信息的传递或交流的过程.用任何方法,通过任何传输媒质将信息从一点传输到另一点的过程称为通信.传统的通信其信息表达方式是语言和文字,信息传输是采用有线电和无线电的方式.1838年,英国建成第一条电报线路.1844年,美国莫尔斯(Morse,S.)建立了莫尔斯电报系统.1876年,美国贝尔(Bell,A.G.)发明了实用的电话系统,从而在19世纪末使许多国家建立了人工交换的有线电话网.1873年,英国麦克斯韦(Maxwell,J.C.)建立了电磁波理论,14年后,德国赫兹(Hertz,H.R.)完成了电磁波辐射实验.在此基础上意大利人马可尼(Marconi,G.)和俄国人波波夫(Попов,A.C.)分别于1895年和1896年各自独立地实现了无线电报.第一个播送语言和音乐的无线广播电台于1919年在英国建成.1929年,英国广播公司进行了最早的电视试播.以上所述的通信方法都属于以传递连续信号为特征的模拟通信范畴.模拟通信与以后发明的数字通信相比较,在长距离上,抗干扰能力较差,不便于进行现代加密.20世纪30年代,美国人奈奎斯特(Nyquist,H.)和苏联科学院院士卡捷尔尼可夫(Котельников,B.A.)分别提出了采样定理,从理论上证明了可以把时间上连续的模拟信号变为离散信号.1937年,提出了脉码调制(PCM),把离散的模拟信号变换成数字信号,为数字通信奠定了基础.

第二次世界大战后,随着电视的普及,微波通信也迅速发展起来.20世纪50年代,一些国家已建成以微波接力通信为主体的通信网,并在不便设立视距接力站的地方建立了超视距散射通信系统.20世纪60年代,数字通信开始发展,它先用于市内电话和长途电话的中继线,以扩充未经复用的模拟通信的容量.1965年,美国发射了第一颗实用静止通信卫星之后,20世纪70年代,容量极大的光纤通信迅速发展起来.同时,随着其他数据数字业务的发展,交换技术进入了程控交换技术的新阶段,从而使得由交换设备、传输设备和用户设备组成的通信网得到了普及,并在综合数字网基础上又向综合业务数字网发展.单向的一般通信系统由信源、信宿、发端设备、收端设备和信道(传输媒质)等五部分组成,如图所示.



信源就是消息的来源.信源发出的消息形式可以是连续的,也可以是离散的.信宿又称受信者,即利用发端信息的对象,它可以是人,也可以是机器设备.发端设备将信源发出的消息转变成适宜在信道中传输的信号.发端设备包括信号的变换、编码、调制及其他附加设备,但也可能不包含编码环节.变换是把各种物理量表示的消息(如声、光信息)变换成电信号.编码是指将模拟信号变为数字信号,常用脉码调制和增量调制.调制是为了使信号在带通型信道中传输,对载波进行调制,通常可分为模拟调制(例如调幅、调频和调相等)和数字调制(例如幅移键控、频移键控和相移键控等).对于基带型信道,经过码型变换处理,使信号达到有效传输.对于光纤通信,需由上述电信号控制一个光源,实现信号的电—光转换.信道也称传输媒质.一般分为有线信道、无线信道和光信道.在有线信道中通信时,电磁波沿线路传播;在无线信道中通信时,电磁波在自由空间(或地下、水下)不定向或定向传播.在光纤通信中,信道就是光纤.信号在传输媒质中传输会受到衰减,出现衰落、失真、引入噪声等.收端设备的作用与发端设备相反,主要完成解调、解码、反变换等任务,对于光纤通信,首先要实现光—电变换,将复原的信号提供给信宿.

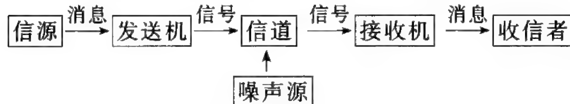
通信数学理论(mathematical theory of communication) 用数学来阐明通信的本质的理论.广义地说,它涉及1928年哈特利(Hartley,R.V.I.)和奈奎斯特(Nyquist,H.)分别提出的信息概念和信息率与频带的关系;还涉及1942年维纳(Wiener,N.)和1956年卡捷尔尼可夫(Котельников,B.A.)各自引用统计观点说明噪声和信号特性.狭义地说,通常是指基于仙农(Shannon,C.E.)于1948年发表的名为"通信的数学理论"的著名论文所形成的信息理论,即信息论.该论文发表在《贝尔实验室技术杂志》第27卷上,全文共五章.文章指出,通信的基本问题是通信双方之一方能够精确地或近似地再生另一方发出的信号,并对此作了定量描述.仙农首先提出信源所发出的消息是离散的或连续的随机过程,并把这种随机性的不定量用信息熵(简称熵)来度量.例如,在信源产生的是单个符号的(最简单的)情况下,熵被定义为

$$H(P_1,P_2,\cdots,P_n)=-\sum_i P(x_i)\lg P(x_i)$$

$$(0\leqslant P(x_i)\leqslant 1,\forall i;\sum_i P(x_i)=1),$$

其中 $P_i=P(x_i)$ $(i=1,2,\cdots,n)$，为信源产生的 n 个符号中的第 i 个 x_i 的概率. 当对数的底数是 2 时，H 的单位为比特. 论文还对信源、信道、编码、译码和信宿的概念作了精确的描述，进而建立了有关通信系统的完整的数学模型. 文章导致了信息论新学科的建立，强调了概率论思想在通信中的地位，并得出信道容量公式、信道编码定理和信源编码定理等一系列重要结果. 其中，著名的仙农公式给出了通信系统信道容量的理论界限. 该论文的发表在科学界和工程界引起了极大的反响，国际权威性杂志《数学评论》立即给予评述. 不到五年的时间，美国电气与电子工程师协会就正式成立了信息论学科组，并于 1954 年出现了信息论汇刊. 目前，通信的数学理论已包括信息论、抗干扰编码理论和信号检测理论，并且随着信息科学的发展而更加完善.

通信系统(communication system) 通信理论研究的基本系统. 指完成通信全过程的所有设备和信道. 通信系统的基本模型如下图所示：



由信源产生的各种待传的消息(语言、文字、图片等)转换为电信号，经发送机将携带消息的电信号变换为适合给定信道中的传输信号，并将其送入信道. 接收机将信道中所传来的带有噪声的信号进行与发送机相反的变换，恢复出相应的电信号，并转换成相应的消息，传给收信者. 模型中的噪声源是代表各设备中以及信道中产生的干扰和噪声的总和.

通信网(communication network) 通信理论术语. 指由多个通信链路把多个通信节点互相连结所构成的一个通信整体. 通信网的功能是适应用户呼叫的要求，实现网中任意两个或多个用户间的信息传输. 通信网包括支撑通信的各种设备，有用户设备、传输设备、交换设备和处理设备等. 为使通信网能够合理协调地工作，通信网还包括网络结构、网络协议. 通信网可从不同角度分类. 按信号形式可分为模拟网和数字网；按业务内容可分为电话网、电报网、传真网、图象传输网、数据通信网等；按通信范围和用途可分为农村网、市内网、长途网及国际网等；按服务对象可分为公用网和专用网. 电报、电话相继问世后，电话网发展相当迅速，以后出现的非电话业务(如传真)也多以公众和专用电话网为基础. 随着计算机的出现和计算机通信的需求，兴起了数据通信，并建立了公众数据网. 目前，新兴的数字网和传统的模拟网已形成了并存和兼容的局面，并逐步向综合业务数字网过渡.

通信网数学模型(mathematical model of communication network) 通信网的一种描述. 是用字母、数字以及符号来表达通信网的组成要素以及它们之间的关系的模型. 通信网往往是一个多目标的决策系统，根据多目标、多层次结构建立数学模型，再根据客观条件进行择优算法处理. 数学模型按用途可分为预测模型、描述模型、优化模型、评价模型，常用的还有电话密度模型和话务量模型等. 因通信网涉及的问题太多，只能把通信网数学模型分成各种特定问题模型分别加以研究. 预测模型的统计调查项目包括用户分布、业务需求、设备性能、技术、费用、地理、资源等，将网络有关数据，用回归分析法或采用数理统计方法来求解；描述模型可针对网的容量、呼损率、时延、可靠性、费用等，采用可靠性理论、排队论、更新论来分析诸因素与指标的关系，如节点之间的最短距离、最大流量途径等问题；优化模型针对多方案的比较，又分为网络模型和数学模型两种，前者使用图的数学模型，后者使用代数方程；评价模型针对备选方案的评比，做出抉择，所用理论有价值理论、决策论、对策论等.

为了规划一个市话局的区域网，就必须对未来用户的数目进行预测，这可以根据人口、电话密度(100 个居民话机占有数)和用户话机随时间增长率等来确定，称此为电话密度模型. 同样，为了规划一个中继网或干线网，要预测每个市话区的始发话务量(即忙时爱尔兰数)，并参考国民生产总值来考虑，称此为话务量模型.

电话密度模型(model of telephone density) 见“通信网数学模型”.

话务量模型(model of traffic) 见“通信网数学模型”.

综合业务数字网(integrated services digital network) 简称 ISDN. 可提供综合服务的数字网. 国际电报电话咨询委员会(CCITT)对于综合业务数字网的定义是：“ISDN 是用同样的数字交换设备和数字传输通道对各种电信业务(例如电话、数据等)建立连结的同一个综合数字网.”简单地说，ISDN 是用一个统一的数字电话网把多种电信业务综合在一起的数字通信网. ISDN 是一个端对端的数字连结的数字网络. 它支持包括话音、数据、文字、图象在内的各种综合业务，向用户提供一种标准的多用途入网接口.

综合业务通信网(integrated services network) 简称 ISN. 一种综合通信网. 能进行各种业务的通信网. 它是可提供电话、电报、数据、传真和电视电话、图象等多种通信服务的网络.

数据通信网(network of data communication) 一种通信网.指传输数据的通信网.通常由计算中心系统、终端子系统和通信子系统组成.由于数据的处理、存储等均由数字计算机进行,所以数据通信网一般均与计算机结合在一起,并在某些点备有计算机.计算机通过数据终端和通信网收集有关各点的数据并加以处理,再根据需要,把处理过的数据向有关点发送出去.有时许多点均有计算机,与数据通信网结合就构成一个计算机网络,实现计算机资源共享.

传真通信网(faxmail network) 一种通信网.即采用传真机和多条通信线路连结成的通信网.它能相互传送各种文字文件、图表、照片等.按用途可分气象、新闻和机要业务等.

电报通信网(telegraph network) 一种通信网.即由数个电路将若干个电报局、用户的电报设备连通,组成能相互进行通报的通信网.可分为公众电报通信网和用户电报通信网两种.公众电报是指发报人将电报送到发端电报局、所,发报局将报文变成电信号经电报通路传送到收端电报局,由收报局、所将电报送到收报人手里.用户电报是指电报用户利用装在用户指定地点的机器设备(通常是电传打字机),通过电报交换机将电路暂时接通,由接通的用户自行通报的一种业务.

电话通信网(telephone network) 一种通信网.即由数个电路将若干个电话局、用户的电话机连通,组成能相互通话的通信网.该网由传输系统、交换系统和信令系统组成.传输通路及其所传信号可以是模拟的,也可以是数字的.一般称前者为模拟电话网,简称电话网;后者被称为数字电话网.

模拟电话网(analog telephone network) 见"电话通信网".

数字电话网(digital telephone network) 见"电话通信网".

专用通信网(private network) 一种通信网.指专供某部门内部业务使用的通信网.专用通信网区别于公用通信网,一般指各部门、机关、企业为本单位业务需要而建设的通信网.例如军事、铁道、水利电力、交通、广播电视、石油等部门的通信网,它专供本部门内部业务使用.公用网是国家电信部门建设、经营的,为整个社会服务的.

国际电信联盟(international telecommunication union) 简称ITU.国际性的电信标准机构.其总部设在瑞士日内瓦,宗旨是在改善和合理运用电信方面维持并增进国际合作,提高电信业务的效率,同时,为普及电信而促进电信技术的发展及其有效的运用,并为实现上述目标而调整各参加国的关系.该联盟至1983年3月止共有158个会员国.中国自1972年起一直担任理事.

国际电报电话咨询委员会(consultative committee of international telegraph telephone) 简称CCITT.有关电报及电话业务、技术等的国际机构.成立于1957年.总部设在瑞士日内瓦,它的职责是对有关电报及电话技术、业务和资费问题进行研究并提出建议.该委员会下设16个研究组.CCITT出版有各研究组的研究报告和建议书等.现已改为ITU-T.

国际无线电咨询委员会(consultative committee of international radio) 简称CCIR.国际电信联盟(ITU)的四个常设机构之一.其任务是对无线电的技术和运用问题进行研究并提出建议.其前身是1927年设立的国际无线电通信技术咨询委员会,在1932年改称现名.该委员会下设13个研究组.其出版物有《国际无线电咨询委员会的建议和报告》,每四年修订一次.现已改为ITU-R.

数字通信

数字通信(digital communication) 一种通信方式.即采用数字化信号传输信息的通信方式,包括数字基带通信和数字载波传输两种形式.在数字通信中的发信端,如果待传输的原始信息是模拟信息,首先对信息进行数字化取样和编码(即信源编码),然后再根据传输信道的要求进行信道编码;如果待传输信息已经是数据信息,则直接进行信道编码.对于数字基带通信,信道中传输的是经信道编码后的信息;对于数字载波通信,经信道编码后的待传输信息对某一载波信号进行调制,信道中传输的是该调制载波.在接收端,经解调、数字化判决和再生后进行信道解码和信源解码,以恢复出原始信息.相对于模拟通信,数字通信有以下特点:

1. 由于采用数字化信号,传输中的信号抗干扰能力大大增强.

2. 数字化信号可以实现中继再生,能够消除传输中引入的各种恶化.

3. 可以大量采用数字电路,简化了系统设备,也易于实现信号交换、多路信号复用.

4. 可以采用数字信号处理技术,有利于实现多种类型业务的传输.

脉码调制(pulse code modulation) 简称PCM.将模拟信号变换为二进制数字信号的一种调制方式.它是将时间连续的模拟信号经过取样、量化、编码而变成脉冲数字信号.其理论根据是奈奎斯特取样定理.脉码调制广泛应用于数字电话通信中.国际电报电话咨询委员会(CCITT)已推荐了两种脉码调制电话传输标准:以2048kb/s速率传输的B扩线A律30/32路TDM-PCM标准和以

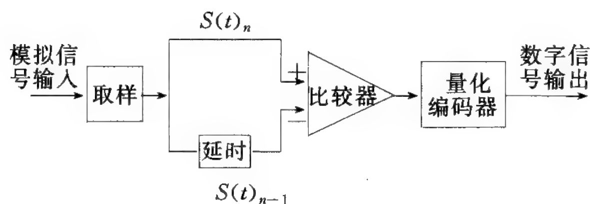
1544kb/s 速率传输的 15 扩线 μ 律 24 路 TDM-PCM 传输标准. 取样频率为 8000 次/秒，每样值用 8bit 编码. 因此，脉码调制后的每个话路码率为 64kb/s.

差分脉码调制(differential pulse code modulation) 简称 DPCM. 脉码调制的一种改进方式. 由于脉码调制是对每一个样值的量化值直接进行编码的，它没有充分利用信号前后幅度样值间的相关性，因而剩余信息较多. 差分脉码调制则是对相邻两个样值的差值量化值编码，而相邻两个样值的变化远小于总体上各个样值的变化，因此，差分脉码调制可用较少的比特数获得与脉码调制同样的信号/量化噪声比.

目前，差分脉码调制编码每个话路 32kb/s 传输速率，可达到脉码调制方式 64kb/s 传输速率的质量. 差分脉码调制基本原理是将上一个样值($S(t)_{n-1}$)保持(延时)一个取样周期，然后与当前的样值($S(t)_n$)进行比较(见下图)，对$S(t)_n$ 与$S(t)_{n-1}$ 的差值

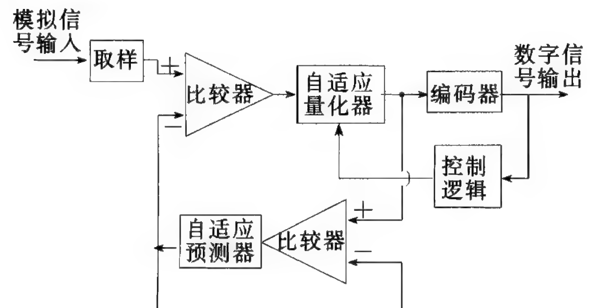
$$\Delta = S(t)_n - S(t)_{n-1}$$

进行量化编码，即得到差分脉码调制信号.



差分脉码调制基本原理框图

自适应差分脉码调制(adaptive differential pulse code modulation) 简称 ADPCM. 差分脉码调制的一种改进方式. 它是为了进一步提高信号/量化噪声比(S/Q). 差分脉码调制预测器的输出值是根据前三个样点的幅度值，采用固定权值的常量调整预测器. 而自适应差分脉码调制预测器则是根据前几个样值的统计特性，采用自适应方式调整输出，并且量化器的量化尺度也随其输入的两个样值的差



自适应差分脉码调制原理框图

的大小进行增、减调整(见图). 因此，自适应差分脉码调制不仅可使数据速率降低(目前一般可比脉码调制速率低一半)，而且可获得更高的 S/Q 值.

增量调制(delta modulation) 简称 DM 或 ΔM. 脉码调制的一种特殊方式. 其基本原理是：对模拟信号的增量进行编码，对当前的样值与前一样值进行比较. 若前者大于后者，编码器便产生一个“1”脉冲；反之产生“0”脉冲. 因此，增量调制产生的脉冲代表该时刻的信号与前一个信号的增量(差值)，而不表示该瞬间信号的样值本身. 由于增量调制的量化阶是固定的，因而存在动态范围小的缺点，不能满足35～40dB 动态范围的要求. 增量调制的每话路16kb/s传输速率可达到中等话音质量的水平.

自适应增量调制(adaptive delta modulation) 简称 ADM. 增量调制的一种改进形式. 其特点是：量化器的量化阶能自动地随信号幅度的大小而变化，从而扩大动态范围. 如果量化阶大小是由直接检测输出数码中的平均斜率信息(在音节 10ms 内的平均值)来控制的，就称为数字检测音节压扩增量调制；如果量化阶的控制取决于相邻三个数码，则称为瞬时压扩增量调制；如果在大信号段采用音节压扩，而在小信号段采用瞬时压扩，则称为混合压扩增量调制；如果量化阶控制信息直接由输入模拟信号中提取，则称为连续增量调制；如果把模拟信号经过积分后再进行音节压扩增量调制，则称为音节压扩总和增量调制，简称音节压扩.

样值(sample) 通信理论术语. 指一个时间连续信号在某一取样瞬间的值.

取样(sampling) 通信理论术语. 指对信号进行抽取样值的过程. 通常指对时间连续信号进行取样. 当取样脉冲为单位冲激脉冲序列时，称为理想取样，或瞬时取样. 当取样脉冲为具有一定宽度的脉冲序列时，称为自然取样. 它的取样值是存在于取样脉冲持续时间内的连续信号值.

模/数变换(analog to digital converting) 简称 A/D 变换. 通信理论术语. 指将时间上连续的模拟信号进行取样、量化编码，成为数字信号的过程.

信源编码(source encoding) 一种编码. 指为提高传输效率而对信源符号进行的编码. 即对信源产生的符号序列进行有效编码，而尽量减少符号序列剩余的一种变换过程.

纠错码(error correcting code) 一种编码. 指在传输过程中发生错误后能在收端自行纠正的编码. 为了使信息码元具有纠错能力，必须对原信息位码元增加冗余位码元，组成一个码字，以增加剩余度. 由信息位与冗余位按某种关系生成有一定剩余度的码字的过程，称为编码. 若信息位与冗余位是线性关系，则称为线性码；若是非线性关系，则称为非

线性码. 将收到的码字用相应监督关系去校验，发现并纠正传输中引入的差错的过程称为译码. 常用的纠错码可分为两类：分组码和卷积码.

检错码(error detecting code) 一种编码. 指在传输过程中发生错误后，在接收端能自动检查并发现错误的编码. 目前常用的检错码有奇偶校验码、恒比码等. 奇偶校验码是在原信息码元后面附加一个监督元，使码组中 1 或 0 的个数为奇数或偶数，为奇数者，称为奇数校验码；为偶数者称为偶数校验码. 奇数校验码可以发现奇数个错误. 恒比码是码组中 0 和 1 的比例相同的编码. 它能检查并发现偶数个错误.

误比特率(bit error rate) 简称 BER. 衡量数字传输系统质量的重要指标. 指所传输、处理的信息发生差错的概率.

误比特率与每比特信号能量、信号波形及噪声功率谱密度有关. 当信号波形给定时，误比特率与信号噪声能量比有关. 在高斯信道下，信噪比越大，误比特率越小.

线性分组码(linear block code) 纠错、检错码的一种. 将信源输出的序列，先进行分组，再对各分组信息源按一线性规律加入冗余位构成码字，该码字称线性分组码.

同步(synchronization) 通信理论术语. 指在数字通信网中，调整两个数字信号的有效瞬间，彼此达到并保持确定的相位、速率关系的过程. 依调整对象和控制机理的不同，同步可分为位(码元)同步、复接同步、帧同步、网同步等.

位同步(bit synchronization) 数字信号码元时间对齐的过程. 指数字传输中，从信息码流中提取时钟信号，并借助于时钟信号来识别信息码元的定时提取过程.

复接同步(multiplexing synchronization) 多路复用码流相位对齐的过程. 把若干个低速码流合成为一个高速码流的过程称为数字复接(用). 复接的各个低速码流必须与相应的高速码流保持确定的相位关系，为提供这种确定相位关系所进行的频率或相位调整，称为复接同步.

帧同步(frame synchronization) 数字帧相位对齐的过程. 在数字通信中，数字信息按帧传送时，从所接收的码流中识别并提取帧定时的过程.

网同步(network synchronization) 数字网的定时同步过程. 在数字网中，所有节点各个交换、传输设备相互同步到一个时钟的速度、相位的过程. 网同步有主从同步、分段主从同步、定时参考分配同步等多种方式.

载波同步(carrier synchronization) 模拟传输系统中，发收两端载波频率和相位相互保持一致的过程.

奈奎斯特速率(Nyquist rate) 在理想低通信道中，前后码元的符号间无码间干扰时符号的极限传输速率. 若理想低通信道的截止频率为 f_N，则 $2f_N$ 为奈奎斯特速率. 满足奈奎斯特第一准则时，信号速率可达到 2Baud/Hz.

码间干扰(intersymbol interference) 一种干扰形式. 由于传输设备及信道的特性不理想而引起的前后码元符号间的相互干扰. 发生码间干扰，可能引起接收码元判决错误，从而产生误码.

概率译码(probabilistic decoding) 对卷积码译码的一种方法. 主要有维特比最大似然译码算法、法诺序列译码算法. 前者适用于对短约束长度的卷积码译码，后者适用于对长约束长度的码译码.

透明性(transparency) 通信理论术语. 指传输信道(永久、半永久、交换连结)对通过的数据流不加改变的传输. 透明又可分为码型透明和比特序列透明等.

相位预均衡(phase pre-equalization) 一种预均衡方法. 指对经载波信道传输的数据流所产生的时延失真，在自动均衡前进行初步的相位时延均衡. 一般是用一个频域均衡器实现.

频域均衡器(frequency-domain equalizer) 一种均衡补偿设备. 用以完成对数据传输信道时延/频率特性失真进行均衡补偿的设备. 它只适用于恒参信道的固定均衡.

时域均衡器(time-domain equalizer) 一种均衡器. 基于消除脉冲信号波形在取样点的符号间干扰所建立的补偿系统称为时域均衡器. 根据其工作方式分为人工手动均衡器和自动均衡器两类.

预置自动均衡器(pre-set automatic equalizer) 一种均衡器. 传送用户数字信息前，根据发送的一组测试信号(称为训练码)，对均衡器进行反复调整，从而达到自动均衡的均衡器.

时间压缩复用(time compression multiplexing) 通信理论术语. 指在一对实线上双向传输时分复用信息的一种方式. 它是将数字脉冲序列进行时间压缩，并按时间分割，交替地进行收发双向传输，从而实现了复用.

回波消除法(echo cancellation technique) 克服二线全双工数字传输造成回波的一种技术. 它的基本原理是：建立对回波波形(参数)的估值，并把它从接收信号中减掉，达到消除回波的目的. 对回波波形的估值可采用横向滤波器. 它能产生一个与回波信号大小相等而极性相反的回波信号，用以抵消接收信号中的回波信号，从而达到无回波干扰接收之目的.

信　　道

信道(communication channel)　信息传输系统的基本组成部分.通常传输信息的通道称为信道.它的任务是以信号的方式传输信息和存储信息.当知道信道的输入信号和输出信号以及它们之间的统计依赖关系后,信道的全部特性就确定了.一般情况,输入和输出信号都是广义的时间连续随机信号,故可用随机过程来描述.当输入和输出的随机变量取离散值的信道称为离散信道;当输入和输出的随机变量取连续值的信道称为连续信道.

电波传播(radio wave propagation)　关于无线电波传播过程的理论.它研究电波与媒质间的相互作用及物理机理,并计算传播过程中各种特性参量.1864年,麦克斯韦(Maxwell,J.C.)建立了麦克斯韦方程,预言电磁波的存在.1887年,赫兹(Hertz,H.R.)用实验的方法证实了电磁波的存在.19世纪以来,波波夫(Попов,A.C.)和马可尼(Marconi,G.)分别独立地进行了电波传播的试验.20世纪,电离层的发现以及电离层形成理论的提出,围绕短波电离层传播开展了广泛研究.在二次世界大战期间,开展了对流层电波传播的研究.无线电波的频谱范围极宽,其波长从几千万米到0.1mm左右.无线电波在介质或介质分界面的影响下,有被折射、反射、散射、绕射和吸收等现象,使电波的特性参量,如幅度、相位、极化、传播方向等发生变化.电波传播方式有:

1. 地波传播.电波传播主要受地球表面的影响.

2. 对流层电波传播.电波传播主要受对流层的影响.

3. 电离层电波传播.电波传播主要受电离层的影响.

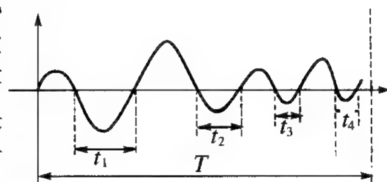
4. 地-电离层波导电波传播.电波传播主要受电离层下缘和地面的影响.

衰落(fading)　通信理论术语.指接收信号电平的随机起伏.即接收信号幅度随时间的不规则变化.起因有多种因素.最主要的起因是多路径传输的干涉、延迟射线随机干涉效应.其次还有传播路径上的降雨衰落、绕射衰落等衰减型衰落和因电波极化方向的变化而引起的极化衰落.它们可表现为时间、频率和空间选择性衰落.根据衰落的速率快慢,通常可分为快衰落和慢衰落.前者是瞬时信号电平在短期内的起伏;后者是短期信号电平中值在长期中的起伏.由于衰落具有随机特性,需用统计方法,如概率密度或分布函数描述.快衰落的幅度分布一般服从瑞利分布.对慢衰落进行较准确的统计分布描述比较难.对流层散射传播中,慢衰落通常服从对数正态分布.衰落特性可用衰落深度、衰落率和衰落持续时间等主要参量来描述.衰落对传输信号的质量和传输可靠性都有很大的影响.严重的衰落甚至会使传输中断.

衰落深度(fading depth)　衰落特性之一.幅度衰落的变化程度.信号电平低于中值的分贝数.即信号电平瞬时值与中值之差,或分布概率分别为50%和10%的电平间分贝数之差.

衰落率(fading rate)　衰落特性之一.即衰落快慢的一个指标.通常指在单位时间内,信号幅度以正斜率通过某给定值的次数.它表明衰落的频繁程度.

衰落持续时间(duration of fading)　衰落特性之一.指信号幅度低于某给定值的持续时间.它是一个随机量,因此需要一个平均量度,即平均衰落持续时间.它是在给定的统计时间内,信号幅度下降到某给定值之下的各段持续时间和的平均值.如图所示,衰落持续时间为$\sum t_i$的平均衰落持续时间为$\sum t_i/T$.



噪声(noise)　通信理论术语.指任一通信系统中传输的有用信号以外的、规则的和非规则的、可懂的和不可懂的干扰.它包括系统内部的噪声(如热噪声、散粒噪声和串话)以及外部干扰(如天电干扰、邻台干扰等).噪声的大小可用等效噪声温度T_e来表示.一个接收系统的总等效噪声温度表示为

$$T_e=\frac{N}{k\cdot B},$$

式中:N为噪声功率;k为玻耳兹曼常数;B为频带宽度(Hz);等效噪声温度用绝对温度(单位为K)表示.

信噪比(signal-to-noise ratio)　衡量噪声对通信影响程度的参量.通信系统中某一端点上的信号平均功率S与噪声平均功率N之比称为信噪比.在数字通信中,通常用归一化的每比特数字信号波形之平均能量E_b与单位频带内的噪声平均功率N_a的比值表示接收端的输入信噪比,以分贝为单位的信噪比可由下式表示:

$$\text{信噪比}=10\lg\frac{S}{N}\ (\text{dB}).$$

多径效应(multipath effect)　由多条路径传播引起的干涉时延效应.因为各条传播路径会随时间变化,参与干涉的各分量场之间的相互关系也就随时间而变化,由此引起合成波场的随机变化,从而形成总的接收场的衰落.因此,多径效应是衰落的重要成因.多径效应对于数字通信、雷达最佳检测等都有

着十分严重的影响.

传输损耗(transmission loss)　传输过程中发生的信号能量损失. 对于某一传输电路,发射天线输入功率与接收天线输出功率之比称为该电路的传输损耗,若用分贝表示则为

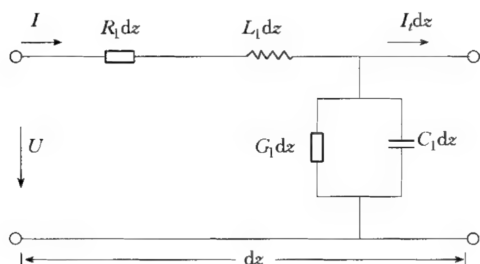
$$L_t = 32.45 + 20\ \lg f + 20\ \lg d + A - G_t - G_r,$$

式中 f 为工作频率(MHz);d 为传输距离(km);A 为馈线衰减(dB);G_t 和 G_r 分别为发射天线与接收天线的增益(dB). 当发射天线、接收天线各向同性($G_t = G_r = 0$)时的传输损耗称为基本传输损耗,用分贝表示为

$$L_b = 32.45 + 20\ \lg f + 20\ \lg d + A.$$

传输线(transmission line)　一种以横电磁(TEM)模的方式传送电能和电信号的波导结构. 其作用是将电磁场的能量集中在它的周围,并将电磁波导向目的地. 传输线的主要结构形式有平行双导线、平行多导线、同轴线、带状线以及工作于准 TEM 模的微带线等. 它们都可借助简单的双导线模型进行电路分析. 各种传输 TE 模、TM 模或其混合模的波导都可以认为是广义的传输线.

传输线方程(transmission-line equation)　阐述传输线上电压 U 和电流 I 之间关系的微分方程



组. 如图为均匀双导线一个微分段 dz 的等效电路. 假定该传输线段上的电压、电流是简谐变化的,简谐信号(简谐因子 $e^{j\omega t}$)的传输角频率为 ω,则传输线 dz 段的串联阻抗为

$$Z_1 dz = (R_1 + j\omega L_1) dz;$$

并联导纳为

$$Y_1 dz = (G_1 + j\omega C_1) dz;$$

传输线方程为

$$\frac{dU}{dz} = Z_1 U(z),$$

$$\frac{dI}{dz} = Y_1 I(z),$$

其中 $U(z)$,$I(z)$都由含因子 $e^{j\omega t \mp U(z)}$ 的两项组成,分别表示朝 $\pm z$ 方向传输的行波(即入射波和反射波).

传播常数(propagation constant)　描述由电压或电流行波沿传输线行进过程中的衰减和相移的参量. 它通常是一个复常数,即

$$\gamma = \alpha + j\beta$$

式中,α 称为衰减常数(Np/m,dB/m);β 称为相移常数(rad/m).

信道编码(channel coding)　一种为了提高通信系统可靠性而对传输信号进行的编码. 即使所传输信号内部具有更强的规律性和相关性,以使信号受干扰而损伤时仍能正确恢复原有信息的处理过程. 即为了适应信道条件(如频带宽度限制、波段、功率、通信时间等)和满足通信的要求(如可靠性、有效性等),需要对所传输的信号进行的某种变换. 信道编码的理论基础是信息论中的信道编码定理,它从理论上解决理想编码器的存在性问题. 信道编码的具体方法属于差错控制技术.

保 密 通 信

保密通信(secret communication)　一种通信方式. 是对要传送的信息内容采取特殊措施,从而隐蔽信息的真实内容,使非法截收者不能理解通信内容的含义. 为了隐蔽通信内容,通常在传输前,先将信息进行各种形式的变换,形成加密信息,在接收端进行相应的逆变换以恢复原信息. 保密通信以其对通信信息的加密方式可分为模拟保密通信和数字保密通信.

模拟保密通信(analog secret communication)　对模拟通信进行保密的一种通信方式. 在模拟保密通信中,通常利用倒频、频段置乱、时段置乱等方法将待传送的模拟信号按频域或时域进行分割、移置、倒置. 模拟保密通信的特点是信号置乱后的带宽基本保持不变.

数字保密通信(digital secret communication)　对数字通信进行保密的一种通信方式. 在数字保密通信中,对数字信号(包括由模拟信号转换而成的数字信号)的加密方法是直接用相同速率的序列密码对该数字信号进行重新组合. 这种序列密码通常是由有限的随机数组成的密钥,它是通过一个算法器产生的序列,为了防止被破译,密钥是要经常更换的,改变了密钥就改变了序列密码. 在收信端,为了恢复原信息,收信端的序列密码必须与发信端的一致,并一一对应. 为此,除了收信端的密钥和算法器必须和发信端完全一致外,收发信端的序列密码还必须同步.

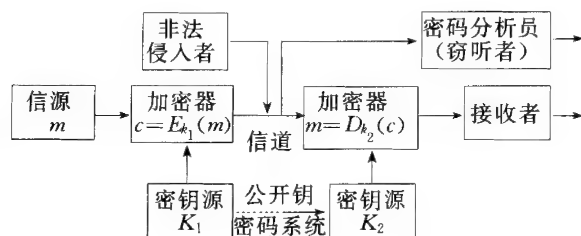
数字保密通信系统(digital secret communication system)　采用数字加密方法实现保密的通信系统. 数字保密通信系统由以下几部分组成:明文消息空间 M;密文消息空间 E;密钥空间 K_1 和 K_2;加密变换 $E_{k_1}: M \to E$,其中 $k_1 \in K_1$;解密变换 $D_{k_2}: E$

$\to M$，其中 $k_2 \in K_2$. 称总体 $(M, E, K_1, K_2, E_{k_1}, D_{k_2})$ 为数字保密通信系统（见框图）. 对于给定明文消息 $m \in M$，密钥 $k_1 \in K_1$，加密将明文 m 变换为密文 c：

$$c = f(m, k_1) = E_{k_1}(m) \quad (m \in M, k_1 \in K_1),$$

合法接收者利用其知道的解密密钥 K_2 对收到的密文进行变换恢复出明文消息 m：

$$m = D_{k_2}(c) \quad (m \in M, k_2 \in K_2).$$



根据加密密钥与解密密钥是否相同，数字保密通信系统分为私钥密码系统（亦称对称密码或单钥密码）和公开钥密码系统（亦称非对称密码或双钥密码）. 私钥密码系统要求加密密钥与解密密钥相同，此时密钥需事先约定或者经过安全的密钥信道由发方传给收方. 公开钥密码系统的加密密钥与解密密钥是不相同的，加密密钥可以在网络上广播或者通过密钥管理中心查找.

网络中的保密通信普遍采用私钥和公开钥密码相结合的混合加密体制，即信息加解密时采用私钥密码，密钥传送则采用公开钥密码. 这样既解决了密钥管理的困难，又解决了加解密速度的问题.

私钥密码系统（private-key cryptosystem） 一种数字保密通信系统. 指加密和解密都使用相同的密钥的保密系统. 此密码体制的安全性就是密钥的安全，如果密钥泄露，则此密码系统便被攻破. 最有影响的私钥密码系统是 1977 年美国国家标准局颁布的数据加密标准(DES). 这种密码系统的优点是安全性高、加解密速度快. 缺点是：随着网络规模的扩大，密钥的管理成为一个难点；无法解决消息确认问题；缺乏自动检测密钥泄露的能力. 如果以密码算法对明文的处理方式为标准，则可将私钥密码系统分为分组密码系统和序列密码系统.

序列密码的加密过程是把明文序列与等长的密钥序列进行逐位模 2 加. 解密过程则是把密文序列与等长的密钥序列进行逐位模 2 加. 序列密码的安全性主要依赖于密钥序列. 序列密码的优点是：处理速度快，实时性好；错误传播小；不存在串破译问题；适用于军事、外交等保密信道. 缺点是：明文扩散性差；插入信息的敏感性差；需要密钥同步.

分组密码的加密方式是首先将明文序列以固定长度进行分组，每一组明文用相同的密钥和加密函数进行运算. 为了减少存储量和提高运算速度，密钥的长度一般不大，因而加密函数的复杂性成为系统安全的关键. 分组密码设计的核心是构造既具有可逆性又有很强的非线性的算法. 加密函数重复地使用代替和置换两种基本的加密变换. 分组密码的优点是：明文信息具有良好的扩散性；对插入信息的敏感性强；不需要密钥同步；较强的适用性，适宜作为加密标准. 缺点是：加密速度较慢；错误扩散和传播强.

公开钥密码系统（public-key cryptosystem） 一种数字保密通信系统. 指加密和解密钥不相同，加密钥可以公开的保密系统. 公开钥密码系统是 1976 年迪费(Diffie, W.)和赫尔曼(Hellman, M. E.)提出的一种新型密码体制. 在公开钥密码系统中，每个人都拥有一对密钥，一个用来加密，称为公开钥，另一个用来解密，称为私钥. 由于通过公开钥不可能得到私钥，所以每个人只要保护自己的私钥，而不需要记忆别人的公开钥. 当 A 给 B 发送秘密信息时，A 通过查询得到 B 的公开钥，然后利用 B 的公开钥加密信息得到密文. 当 B 收到 A 的密文后用自己的私钥解密得到明文信息.

公开钥密码系统的安全性基于一些数学问题的难解性. 代表性的有三类：大整数的分解、有限域上的离散对数和椭圆曲线上的离散对数. 最有名的公开钥密码系统是 1977 年由瑞费斯特(Rivest)、沙米尔(Shamir)和阿德尔曼(Adleman)三人提出的 RSA 密码体制，它的安全性基于大整数的难分解性. RSA 密码体制是使用最广泛的公开钥密码系统. 基于椭圆曲线的公开钥密码系统非常适合于智能卡中. 公开钥密码系统的缺点是加解密速度慢，不适合大数据量的加密.

数据加密标准（data encrypting standard） 简称 DES. 进行数据加密保护的标准算法. 在所有分组密码系统中，数据加密标准毫无疑问是最著名的，它的产生被认为是 20 世纪 70 年代信息加密技术发展史上的两大里程碑之一.

美国国家标准局于 1973 年 5 月发出通告，公开征求一种用于对计算机数据在传输和存储期间进行加密保护的标准算法. 1975 年，美国国家标准局接受了国际商业机器公司(IBM)推荐的一种密码算法，并向全国征求采用该算法作为美国信息加密标准的意见. 经过两年的激烈争论，美国国家标准局于 1977 年 7 月正式把该算法作为美国数据加密标准. 1980 年 12 月，美国国家标准局正式采用这个算法作为美国的商用加密算法.

DES 是一种按分组方式工作的对称密码算法. 其基本思想是：将二进制的明文序列分成 64 比特块，用长为 56 比特的密钥对其进行 16 轮代换和换位加密，最后形成密文. 除了密钥输入顺序之外，

DES的加密和解密步骤完全相同，这就使得在制作DES芯片时，易于做到标准化和通用化，这尤其适合现代通信的需要. 在DES出现以后，经过许多专家学者的分析论证，证明它是一种性能良好的数据加密算法，不仅随机复杂度高，而且易于实现，因此，DES在国际上得到了广泛的应用.

但是，20世纪90年代在DES的破译方面取得了突破性的进展. 破译者能够用穷举法借助网络计算在短短的20余小时就攻破56位的DES，所以DES在当代已经是不安全的加密算法了.

高级加密标准(advanced encrypting standard) 简称AES. 一种比数据加密标准(DES)更安全的数据加密标准算法. 随着计算能力的不断提高，DES的加密强度越来越不能满足人们的要求. 在这种情况下，美国国家标准局在1997年4月15日开始倡导制定高级加密标准替代DES以满足21世纪的信息加密需求，并为此成立了AES工作小组. 经过几年的招标、筛选，美国国家标准局于2000年底最终确定了AES——Rijndael. 高级加密标准是比利时的两个密码专家爵德曼(Joan Daemen)和文斯特雷曼(Vincent Rijmen)共同设计的. 高级加密标准的设计策略是轨迹策略. 这种策略是针对差分分析和线性分析提出的. Rijndael是一个迭代分组密码，其数据分组长度和密钥长度可以为128/192/256位.

高级加密标准的加密与解密过程在形式上是一样的，涉及的基本运算单元是8比特字节和32比特字，因此特别适合智能卡和不同平台的软件、硬件实现.

数字签名(digital signature) 保密通信术语. 指采用数字的方法实现电子文档的辨认和验证的签名. 它是传统文件手写签名的模拟，能够实现用户对电子形式存放消息的认证.

数字签名同传统的手写签名相比有许多特点. 首先，在数字签名中签名同消息是分开的，需要一种方法将签名与消息绑定在一起，而在传统的手写签名中，签名认为是被签名消息的一部分；其次，在签名验证的方法上，数字签名利用一种公开的方法对签名进行验证，任何人都可以对签名进行验证，而传统手写签名的验证是由经验丰富的消息接受者通过同以前的签名相比较进行的；在数字签名中，有效签名的副本同样是有效的签名，而在传统的手写签名中，签名的副本是无效的.

数字签名方案一般包括三个过程：系统的初始化过程、签名产生过程和签名验证过程. 系统的初始化过程中要产生数字签名方案中用到的一切参数，有公开的，也有秘密的. 在签名产生过程中，用户利用给定的算法对消息产生签名，这种签名过程可以公开也可以不公开. 在签名验证过程中，验证者利用公开验证方法对给定消息的签名进行验证，得出签名的有效性.

数字签名采取公开钥密码算法或者HASH函数的方法. 当采用公开钥密码算法时，签名者用自己的私钥加密信息，验证者用签名者的公开钥验证. 当采用HASH函数时，签名者计算信息的HASH值，并把信息与该HASH值同时发送；验证者只要再计算一个信息的HASH值，比较HASH值是否相同，就可以确定签名是否有效.

信息伪装(information hiding) 亦称信息隐藏. 保密通信术语. 就是将机密资料隐藏于另一非机密文件中，其形式可为任何一种数字媒体，如图象、声音、视频或一般的文档等. 其首要目标是要使加入隐藏信息的目标媒体产生的最小可见性降低，使人们无法看到和听到隐藏的数据，达到令人难以察觉的目的.

信息伪装必须考虑隐藏的信息在经历各种环境、操作之后而免遭破坏. 在图象传递过程中，机密资料若在非法拦截者破解的过程中消失，则机密资料可以说是安全的；若机密资料是在正常的处理传递过程中消失，则机密资料的传递是失败的、无效的，必须重新传递. 一般的多媒体形式的文件，其数据量都非常大，为了节省传递时间，在传递之前都会将传递的信息资料进行压缩处理. 因此，信息伪装还必须考虑诸如图象压缩这种非恶意操作造成的威胁，使机密资料对正常的有失真压缩技术具有免疫能力. 要求隐藏的数据量与隐藏的免疫力是一对矛盾，不存在一种完全满足这两种要求的方法. 通常只能根据需求的不同有所侧重，采取某种妥协. 信息伪装具有以下几个特性：

1. 隐蔽性. 这是信息伪装的基本要求，经过一系列隐藏处理的图象(宿主图象)没有明显的质量降低，隐藏的信息无法人为地看见或听见.

2. 隐藏场所的安全性. 应将欲隐藏的信息藏在目标图象的内容之中，而非文件头等处，这样可以防止因格式变换而遭到破坏.

3. 免疫性. 隐藏处理后的图象抗拒因图象文件的某种改动而导致隐藏信息的丢失的能力. 所谓改动包括：传输过程中的信息噪声、过滤操作、重新采样、编码、有损压缩、模/数转换.

4. 编码纠错性. 为了保证隐藏数据的完整性，使其经过各种操作和变换后仍能很好地恢复.

5. 隐藏数据的自产生性. 经过一些变换，可能使原图象产生较大的破坏，如果仍能从留下的数据中恢复隐藏信号，而且恢复过程不需要宿主图象，这就要求隐藏的数据必须具有某种自相似特性.

目前，信息伪装的主要研究方法有：拼凑物方法；纹理影射方法；几何不变方法.

信息隐藏(information hiding) 即"信息伪装".

频带倒置(frequency inversion) 模拟保密通信频域加密的方法之一.它是用倒频器将模拟信号的频谱整个倒过来,使高频变低频,低频变高频.实现频带倒置的基本电路是平衡调制器和带通滤波器.平衡调制器可以搬移和倒置频谱,而滤波器可以提取需要的倒置频谱成分.

频段置乱(frequency scramble) 模拟保密通信频域加密的一种方法.它是在频带上用滤波器将模拟信号分成若干个分频段,打乱这些分频段在频域上的相互位置,并将其中若干分频段进行倒频.在分为 n 个分频段的情况下,共有 $n! \times 2^n$ 个不同的频段排列.收信端做相应的逆变换,则可以恢复出原始信号.为了提高破译的难度,即增加置乱度,现代的频段置乱器一般都通过序列密码的控制,不断改变频段置乱的排列.

时段置乱(time scramble) 模拟保密通信时域加密的一种方法.它是在时间上将信号分段,使若干个时段组成一帧,然后打乱一帧内时段的先后次序.发信端要在存满一帧信号后才能按置乱的次序输出加密信号,而收信端则要在存满一帧置乱信号后才能按原次序恢复信号.因此一帧的时延不宜太长,否则将影响正常的通信.但是对于话音信号来说,一帧的时延也不宜太短(即短于一个话音音节),因为在同一音节内的时段置乱不能有效地掩盖话音.通常,时段置乱器的一帧由 8~16 个时段组成,时段长为 20~60ms,一帧的时间在 300~600ms 之间.

无线通信

无线通信(wireless communication) 一种通信方式.即不用导线而利用电磁波的辐射和空间传播来传送信息的通信方式.包括:短波通信、超短波通信、微波通信、红外通信,以及移动通信、卫星通信等.用于提供固定通信和移动通信服务.1831 年,英国的法拉第(Faraday,M.)发现电流可以产生磁场.1873 年,英国的麦克斯韦(Maxwell,J. C.)从理论上预言了电磁波的存在,并证明它是以光速传播的.1887 年,德国的赫兹(Hertz,H. R.)用实验的方法证明了电磁波的存在,证实了麦克斯韦的电磁场理论,从而为无线电波的应用奠定了基础.受赫兹电磁波实验的启发,1895 年,意大利的马可尼(Marconi,G.)进行无线通信实验,实现了接收莫尔斯电报码的通信.同年,俄国的波波夫(Попов,A. C.)也独立地进行了用电磁波传送莫尔斯电报码的实验.在 20 世纪 10 年代,用长波和中波波段实现了越洋无线通信.在 20 世纪 20 年代,用短波实现了长距离的无线通信.在 20 世纪 30 年代,开始用超短波实现无线通信.在 20 世纪 40 年代,发现微波,并开始发展微波接力通信.在 20 世纪 50 年代,提出并实现了对流层散射超远距离通信.在 20 世纪 60 年代中期,利用人造卫星通信,成为跨洋的洲际通信和国际通信的重要手段.从 20 世纪 70 年代至 20 世纪 80 年代,推出了无线移动通信系统.在 20 世纪 90 年代,又推出了数字无线移动通信系统.在 20 世纪 90 年代末,利用中、低轨道卫星构成了覆盖全球的卫星移动通信系统.中国于 1899 年在广州、马口、前山、威远各要塞及舰船上设置了无线电台.1912 年在张家口、武汉、吴淞、福州、广州开通无线电报.1923 年中国建立喀什噶尔电台,与印度通报.1930 年建成上海国际电台,开通与旧金山、柏林、巴黎直达电报.在 20 世纪 50 年代,无线通信得到迅速发展,开通国际、国内电报和电话等多种业务.在 20 世纪 60 年代,开始发展微波通信.在 20 世纪 70 年代开始建立卫星通信,并于 1984 年发射了中国第一颗试验通信卫星.

电磁波以光速向四周辐射.它以直线形式在均匀介质中传播,遇到不同介质或障碍物会产生反射、吸收、折射、绕射或极化偏转等现象.电磁波包括无线电波和光波.利用无固定约束的无线电波传送信息的通信方式,称为无线通信.无线电波的波长不同,其传播特性也不同,因此,可根据波长将无线电波划分为不同的波段,如下表所示.

无线电波段和频段名称表

波段名称		波长范围	频段名称	频率范围
长波	极长波	$(100\sim10)\times10^6$m	极低频(ELF)	3~30Hz
	超长波	$(10\sim1)\times10^6$m	超低频(SLF)	30~300Hz
	特长波	1000~100km	特低频(ULF)	300~3000Hz
	甚长波	100~10km	甚低频(VLF)	3~30kHz
	长波	10~1km	低频(LF)	30~300kHz
	中波	1000~100m	中频(MF)	300~3000kHz
	短波	100~10m	高频(HF)	3~30MHz
	超短波(米波)	10~1m	甚高频(VHF)	30~300MHz
微波	分米波	1~0.1m	特高频(UHF)	300~3000MHz
	厘米波	10~1cm	超高频(SHF)	3~30GHz
	毫米波	10~1mm	极高频(EHF)	30~300GHz
	亚毫米波	1~0.1mm	至高频(THF)	300~3000GHz

利用光波传送信息的通信方式称为光通信.利用红外光(红外线)传送信息的通信方式称为红外线通信.利用激光传送信息的通信方式称为激光通信.无线通信的原理是:在发送端将用户信息加载到快速振荡的电信号上(调制),生成载有用户信息的可

以辐射的电磁波(已调波),利用电磁波以光速传播的特性,将用户信息快速传送到远方;在接收端将收到的电磁波(已调波)变换成电信号,并将所载的用户信息取出(解调)送给受信者.其中,加载的过程称调制,调制后的无线电波称已调波;从已调波中提取还原信息的过程称反调制或解调.无线电波传播的媒质称信道.无线通信系统是由发送端的发射机和发射天线、无线传输信道、接收端的接收天线和接收机组成.发射机完成调制功能,发射天线将已调波以无线电波形式辐射出去.接收天线将无线电波转换为电信号,接收机完成解调功能.发射机一般由振荡器、放大器、调制器、变频器和功率放大器等组成.发射天线必须与发射机的输出阻抗相匹配,使无线电波能以最高的效率辐射到空间.接收机一般由前置放大器、变频器、本地振荡器、中频放大器、解调器和低频放大器等组成.无线通信按所用的无线电波的波段不同,可分为甚长波通信、长波通信、中波通信、短波通信、超短波通信和微波通信等.按用户是否移动可分为固定通信和移动通信.按接力站是否在地面可分为卫星(中继)通信和地面(中继)通信.无线通信有如下特点:

1.开放性.无线电波在空间传播,使无线通信信道具有开放性.优点是特别适用于覆盖区域性或全球性的广播信道通信,如卫星通信、移动通信等;缺点是易于被截获(窃听).

2.灵活性.无线电波的空间传播,使无线通信网的网络构成具有灵活性.适用于网络拓扑、网络节点多变的网络,如卫星网、短波网、野战移动网等.

3.易干扰性.无线电波的空间传播,使无线通信易受外界干扰,包括自然干扰和人为干扰.自然干扰,如太阳黑子、磁暴、电离层骚动、雷电等对短波通信的影响;气候、环境、地形、地物对超短波通信的影响;大气折射、地面反射、雨雷云雾等对微波通信的影响.人为干扰,如工业性干扰、系统间干扰、核爆炸、敌方蓄意干扰等.

因此,无线通信技术的特点在于:开放性引起的多址连结技术;灵活性带来的网络管理技术;易干扰性引发的抗衰落、抗多径、抗干扰、抗截获(保密)、抗摧毁技术.

无线通信已成为可靠、高效、灵活的通信手段,广泛应用于民用和军事通信中,特别是在移动通信中,无线通信技术是惟一的而不可替代的通信手段.

长波通信(long-wave communication)　无线电通信的一种方式.即利用波长大于1km(频率低于300kHz)的电磁波传送信息的通信方式.长波通信可细分为长波、甚长波、特长波、超长波、极长波通信.波长越长,传播衰耗越小,穿透海水和土壤的能力也越强,但相应的大气噪声也越大.随着工作波长的增大,其发射天线尺寸也越大,超长波发射天线的长度可达数十甚至上百公里.长波波长为1~10km,主要沿地球表面以地波传播,陆地传播距离为几十至几百公里,海面传播距离可达数百至数千公里,长波通信只能传电报或低速数据.甚长波波长为10~100km,主要靠大地与电离层(低层)形成的波导进行传播,传播距离可达数千公里乃至覆盖全球.甚长波穿透海水能力较强,适用于水下通信.甚长波通信只能传电报或几十波特的低速数据.超长波波长为1000~10000km,其传播衰减很小,穿透海水能力很强,可深达100m以上,用于远距离和大深度下航行的战略导弹核舰艇间通信.超长波通信传输速率很低,1个码元长达30多秒.极长波波长大于10万公里,其在陆地和海水中的传播衰减都很小,适宜数百米水下的舰艇间通信.极长波通信虽然用常规手段很难实现,但是很有应用前景.

中波通信(middle-wave communication)　无线电通信的一种方式.即利用波长100~1000m(频率为300~3000kHz)的电磁波传送信息的通信方式.中波在白天主要靠地波传播,在夜间则增加了由电离层(E层)反射的天波传播.中波被用于广播和导航.

短波通信(short-wave communication)　无线电通信的一种方式.即主要靠电离层反射的天波传播实现远距离通信.通信受电离层变化的影响十分严重.但随着对电离层变化规律的了解以及对短波信道的频率自适应技术的发展,短波通信质量有了很大提高.由于短波通信具有设备较简单、机动性大,特别是电离层的不可摧毁性,使得短波通信在军事通信、应急通信中得到发展.

超短波通信主要以直线方式传播,考虑地球曲率,其直线传播距离等于视距,视距传输距离约为50km.超短波具有较宽的频谱,多用于移动通信和电视广播.

微波通信(microwave communication)　无线电通信的一种方式.是利用微波段波长的无线电波进行的通信.空间微波以直线方式传播,地面传输距离与超短波相同,约为50km.更长距离通信时需设接力站,构成微波接力通信.用卫星作接力站则构成卫星通信.微波也可沿波导传播.微波波段具有极宽的频谱,故微波通信可以提供更大容量的传输能力.

移动通信(mobile communication)　无线电通信的一种方式.指通信双方或一方处在运动中实现的通信,如移动体与固定点之间、移动体与移动体之间的通信.移动通信按活动范围分为陆、海、空三类;目前按使用情况可分为峰房移动通信、集群移动通信、无线电寻呼、无绳电话等,正朝着个人通信和全球移动通信方向发展.移动通信的发展大体可分为

五个阶段：

1. 20世纪20年代至40年代为早期阶段. 在这一阶段，首先在短波频段上建立了专用移动通信系统，在20世纪40年代，将频段提高到30～40MHz.

2. 20世纪40年代至60年代初为初期阶段. 公用移动通信业务开始问世，并完成了人工无线、有线转换系统，解决了移动电话系统与公用电话网之间的接续问题.

3. 20世纪60年代至70年代中期为改进完善阶段. 使用150MHz和450MHz频段，采用大区制，实现了无线信道的自动选择以及自动接续到公用电话网.

4. 20世纪70年代到80年代中期为移动通信大发展时期. 以蜂房小区制为特征的蜂房模拟移动通信系统商用化，它标志着实用化移动通信网的建立.

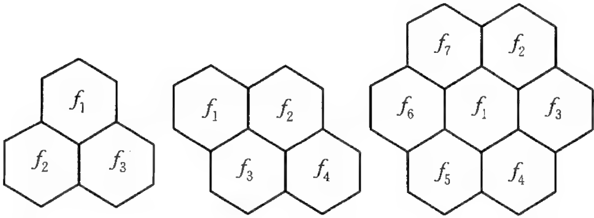
5. 20世纪80年代中期以来，是移动通信从模拟体制向数字体制发展的时期，即由第一代模拟移动通信走向第二代数字移动通信. 具有代表性的有泛欧全球通(GSM)标准下的时分多址(TDMA)数字蜂房体制和美国IS-95标准下的码分多址(CDMA)数字蜂房体制.

目前，移动通信正向第三代迈进，即个人通信. 未来的移动通信将提供全球性优质服务，真正实现在任何时间、任何地点向任何人提供多种业务的通信服务.

大区制(large coverage system)　移动通信中无线电覆盖划分的一种体制. 它是指由一个基站覆盖一个较大的服务区，其覆盖范围仅受限于电波传播损耗，在甚高频段(VHF)和特高频段(UHF)通常大区的覆盖半径约为25～40km. 大区制移动通信其网络结构简单，没有频率重复使用的问题，它适用于用户密度不大或通信容量较小的系统.

中区制(middle coverage system)　移动通信中无线电覆盖划分的体制之一. 其每个无线基站的服务半径比大区制要小，约15～30km，工作方式与小区制基本相同，容量一般在1000～10000用户之间，适用于中容量移动通信系统.

小区制(small coverage system)　移动通信中无线电覆盖划分的体制之一. 它是将一个大服务区划分为许多个半径约2～10km的小覆盖区(小区或称基站区)，每个小区设一无线基站，使用一组频率. 若干个小区组成一个小区群，小区群内的各小区以不同的频率(或相位)划分，而小区群之间频率(或相位)可重复使用. 目前，小区构成的几何图形主要为正六角形，如图所示. 如此构成的形状与蜂房相似，故称为蜂房式小区结构. 由于小区制蜂房结构中，不同小区群之间的对应小区频率可重复使用，因而提



高了频率的利用率，在给定频率资源下，用户容量可大为增加. 因此，小区制适用于大容量的移动通信系统. 小区制成为目前公众移动通信网普遍使用的一种覆盖划分的方式. 而且随着用户容量的剧增，又出现了微小区划分方式.

无线电寻呼系统(radio paging system)　亦称袖珍铃系统. 一种单频、单工、单向无线电通信系统. 它向持有寻呼机的用户发送单向简短的信息. 该系统由一个寻呼中心、寻呼发射机、寻呼控制台及用户寻呼机组成. 其工作过程是：当要寻呼持有寻呼机的用户时，通过公用电话网拨通无线寻呼控制中心，告知欲寻呼的号码和本人电话号码，及欲传的简短信息内容. 寻呼发射机将寻呼的全部信息发送出去. 被叫无线寻呼用户收到寻呼信息后，发出声响，通知用户，并将信息显示在小屏幕上，这次寻呼即结束. 无线寻呼系统按用户类别可分为公用和专用寻呼系统. 公用系统对公众开放，并与公用电话网相连结；专用系统仅供某一单位使用. 寻呼机的显示方式有三种：

1. 用显示一组数字来代表所传的内容方式，称为数显方式.

2. 显示汉字信息的方式称为汉显方式.

3. 显示非汉语文字的显示方式.

寻呼机收到信息时，提醒用户也分声音提示、振动提示两种方式.

目前，随着技术的发展，高速寻呼、双向寻呼等新业务也逐步进入商用.

集群系统(trunking system)　一种专用的模拟无线调度电话系统. 其特点是：多个无线信道共用；中心台(主台)与移动台(属台)为星形网络拓扑结构；使用大区制覆盖；具有通话定时(时间限制)、优先权、组呼、选呼及动态重组功能. 在调度员中心台控制下，移动台之间可相互通话，也可接入公用电话网. 常用于运输调度(货车及出租车调度等)、公安调度、消防调度等调度指挥系统.

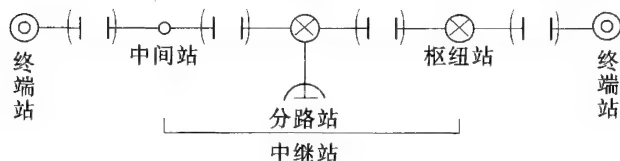
公用移动电话系统(public mobile telephone system)　公用电话网的一个组成部分. 它是无线用户与无线用户之间、无线用户与公用电话网用户之间通过无线基站及无线交换设备和公用交换设备进行通话的通信系统. 公用移动电话网是由小区制的蜂房移动通信系统构成的. 中国蜂房移动通信系统

已由第一代的模拟系统（TACS 体制）向第二代的数字蜂房系统过渡. 数字蜂房体制主要有欧洲的基于全球通（GSM）标准下的时分多址（TDMA）数字蜂房体制和美国 IS-95 标准下的码分多址（CDMA）数字蜂房体制.

短波电台（short-wave radio station） 无线电通信的一种设备. 即由发射机、接收机、天线等组成的短波收发信系统. 发射机是短波信号的发射装置，它将所要传送的话音信号（或其他形式的信号）调制到 3～30MHz 的射频信号上，经放大送入天线，以电磁波的形式发送出去. 接收机从天线接收到空间的电磁波，经过选频、放大、解调，恢复出原调制信号，再经换能设备转换成相应的话音或其他形式的信号. 小型便携式短波电台，发射功率较小，一般采用鞭状天线，主要用于近距离话音通信；中型短波电台传播距离较远，用作电报、数据和话音通信；大型短波电台，一般用于远距离的国际通信.

微波中继通信（microwave relay communication） 长途通信的一种方式. 即利用波长为 0.01～1m 无线电波传送电信号的一种远距离通信方式. 微波在自由空间是以空间波的形式直线传播的，传播的距离因受地球曲率的影响而限制在视距内. 为了延长传播距离，一般每隔 50km 左右（以 46km 为一个标准站）设立一个接力站（即中继站），将接收到的微弱信号进行放大后再重新转发. 中继站的数目依通信距离而定，用中继的方法可使微波通信的距离延长至几千公里.

一个微波中继通信系统的组成示意图如下图，



它由两端的终端站和线路中的许多中继站所组成. 终端站的作用是：在发送端，把基带电信号进行中频调制，然后再变换到微波频率并发射出去；在接收端，将接收到的微波信号进行放大解调，恢复成基带信号输出. 中间站、分路站及枢纽站统称为中继站，它们对微波信号进行接力、转发或分路. 微波中继通信的主要特点是：

1. 由于工作频率在微波段，其通信频带很宽，可用于传输大容量、宽频带的信号，如多路话音及电视信号.

2. 受天电干扰和工业干扰的影响很小，并基本不受昼夜及季节的影响，工作稳定.

3. 天线有很强的方向性且增益高，可以用小功率的发射机而获得好的通信效果.

4. 与电缆通信相比，设备投资成本低，施工安装周期较短，并能克服一些地形带来的不便.

微波中继通信由于有上述特点而成为现代通信的一种主要手段，它主要用于解决城市之间、地区之间、部门之间的大容量信息的传输问题，如多路长途电话、电视节目的传输等. 目前微波通信正从传统的模拟方式迅速地向数字方式发展，数字微波中继通信已得到了广泛应用.

地球站（earth station） 亦称地面站. 一种设置在地面上用以进行空间通信的设备. 它是沟通卫星和连结用户的中枢. 一个卫星通信系统包括许多地球站，其规模从天线直径达 30m 的固定站到天线直径仅为 1m 左右的车载站、船载站等. 地球站一般由天线系统、发射系统、接收系统、跟踪伺服系统和电源系统组成. 固定站多采用大天线、大功率发射机，具有致冷参量放大器的低噪声接收机，以及大容量多路通信设备. 近年来，向小型化（膝上、手持及个人）方向发展，由固定地面站向移动个人终端发展.

卫星通信（satellite communication） 无线电通信的一种方式. 即利用人造地球卫星作中继站来转发无线电信号，在两个或多个地球站之间进行的通信. 卫星通信系统主要由通信卫星、地球站和卫星通信链路组成. 卫星通信的特点是：

1. 通信距离远，能进行越洋和洲际通信. 利用多卫星系统，可实现全球通信.

2. 通信容量大. 卫星通信使用微波波段，具有很宽的频带，一个转发器可传输 12 路电视或几千路电话.

3. 不受大气层骚动的影响，通信可靠.

单频单工（single-channel simplex） 无线电通信中的一种工作方式. 通信双方收发使用同一频道进行单向只发不收或只收不发的通信方式. 单频单工方式的电台设备比较简单，所有用户都使用同一频率，可以直接通话，但一对用户通话时，所有使用这一频率的其余用户均可收到.

双频单工（double-channel simplex） 亦称为异频单工或准双工. 无线电通信中的一种工作方式. 通信双方收发使用两个频道的单工通信.

双频双工（double-channel duplex） 无线电通信中的一种工作方式. 通信双方使用一对频道，且为双工工作方式的通信.

双工（duplex） 无线电通信中的一种工作方式. 通信双方可以同时进行发送和接收的工作方式.

单工（simplex） 无线电通信中的一种工作方式. 通信双方只能交替地进行发送和接收的工作方式. 通常在通信机的送/受话器上安有按键开关，按下开关，通信机处于发送信息状态，松开开关，通信机处于接收状态.

超短波通信(ultrashort wave communication) 无线电通信的一种方式.即利用波长在1～10m左右(频率在30～300MHz)的电磁波进行的无线电通信.超短波通信的传播路径主要靠空间波进行视距范围内的传播,当要进行远距离通信时,需加设中继站进行接力.

流星余迹通信(meteoric trail communication) 无线电通信的一种方式.即利用流星余迹这种不均匀媒质散射无线电波进行远距离无线电通信的方式.流星进入大气层的高速运动产生光和热,把80～140km高空的气体电离成带正电和负电的微粒,形成一条云雾状长带,即流星余迹,在短时间内正负电离子复合而余迹消失(据测,每天约有80亿个流星堕落于大气层中).

流星余迹通信的方法是:把要发送的信息预先放在发信机的存贮器中,当合适的流星余迹一出现,马上由发射机高速发出信息,依靠流星余迹的散射作用,将电磁波散射到接收机,流星余迹一消失,发射机立即停止发射,并等待下一次合适的流星余迹出现.由于流星余迹的存在时间短,故需采取快速通信的方式,常用的频段为30～100MHz.

流星余迹通信的特点是:通信稳定性好,用小功率可进行远距离通信,500W功率发射机的通信距离可达2000km左右.由于电波在空中停留的时间短,方向性强,因而保密性高,在军用通信中起主要作用.不足的是发送状态是断续的,不适于进行实时通信,而且传送数字信号时误码率较高.

电离层散射通信(ionospheric scatter communication) 无线电通信的一种方式.即利用电离层对电磁波的散射作用进行的远距离无线电通信.电离层是指地球表面60～2000km的高空.在电离层中,正负电离子受太阳黑子、地球磁场等各种因素的影响而分布不均匀,形成一团团云状气团,无线电波射入这些不均匀体,发生散射.电离层散射通信所用的工作频段一般为35～70MHz.它除具有对流层散射通信的优点外,还具有设备较简单、灵活,通信距离远(一般可达1000～2000km),可靠性较高等优点,但传输频带很窄,通信容量小,一般只传输一路话音.

散射通信(scatter communication) 无线电通信的一种方式.即利用空中不均匀媒质对电磁波的散射作用进行远距离的无线电通信方式.空气中有许多微小尘埃,使空气成为不均匀介质,无线电波在空中传播时,一小部分波遇到不均匀介质,将改变原传播方向,向四面八方散射,远处的接收机接收这些散射来的微弱电磁波,从而实现远距离通信.散射通信一般使用超短波和微波频段.通信距离最远可达上千公里,但要使用大功率发射机和高灵敏度的接收机以及定向天线等.散射通信包括对流层散射通信、电离层散射通信和流星余迹通信等.

多址连结(multiple access) 建立信道的一种方法.是卫星通信中多个地球站通过共同的通信卫星同时在各地球站址之间建立信道的方法.这一方法在移动通信中也被广泛采用.多址连结的基础是信号分割理论,利用信号的正交性,各个地球站对接收的信号具有识别能力,并能从混合的信号中选择出本站所需的信号,而建立正确信道连结.目前,卫星通信和移动通信中所采用的多址方式主要有频分多址、时分多址、空分多址、码分多址.

频分多址(frequency division multiple access) 简称FDMA.利用频域正交信号实现多址通信的方法之一.是基于信号在频域上的正交性,即对各地球站所发信号在频率域上进行分割,使各地球站信号$x_i(f)$在卫星总频带W内所占的频带Δf_i互不重叠.频率正交性可表示为

$$\int_{\Delta f_i} x_i(f)\cdot x_j(f)\mathrm{d}f=\begin{cases}1 & (i=j),\\ 0 & (i\neq j)\end{cases}$$
$$(i,j=1,2,\cdots,k),$$

式中x_i,x_j分别代表第i站和第j站发送的信号.各地球站通过频率选择(用滤波法),就可以从混合信号中选出所需的站址的信号.

时分多址(time division multiple access) 简称TDMA.利用时域正交信号实现多址通信的方法之一.基于信号在时域上的正交性,即对各地球站所发信号在时间域上进行分割,使各地球站信号$x_i(t)$在时间T内所占据的时隙ΔT_i互不重叠.时间正交性可表示为

$$\int_{\Delta T_i} x_i(t)\cdot x_j(t)\mathrm{d}t=\begin{cases}1 & (i=j),\\ 0 & (i\neq j)\end{cases}$$
$$(i,j=1,2,\cdots,k).$$

各地球站通过时间选择门,就可以从接收的信号中选取出所需站址的信号.

空分多址(space division multiple access) 简称SDMA.利用空间域正交信号实现多址通信的方法之一.基于信号在空间上的正交性,即对各地球站所发信号在空间上进行分割,使各地球站信号处在卫星天线的不同波束覆盖的小空间Δs_i.空间正交性可表示为

$$\int_{\Delta s_i} x_i(s)\cdot x_j(s)\mathrm{d}s=\begin{cases}1 & (i=j),\\ 0 & (i\neq j)\end{cases}$$
$$(i,j=1,2,\cdots,k).$$

空分多址通常与时分多址(TDMA)相结合使用.各地球站处在不同的波束中,通过选择TDMA的时隙,就可以从接收的信号中选出所需站址的信号.

码分多址(code division multiple access) 简

称 CDMA. 利用码序列正交信号实现多址通信的方法之一. 基于信号的正交性,即对各地球站所发信号进行地址码型分割,即每个地球站发送信号的码型相互正交. 码型正交性可表示为

$$\int_T C_i(t)\cdot C_j(t)\mathrm{d}t = \begin{cases}1 & (i=j),\\ 0 & (i\neq j)\end{cases}$$

$$(i,j=1,2,\cdots,k),$$

式中 $C_i(t)$,$C_j(t)$ 是指第 i,j 站的地址码,准正交情况:

$$\int_T C_i(t)\cdot C_j(t)\mathrm{d}t \approx 0 \quad (i\neq j).$$

各地球站通过对地址码的相关检测接收就可以从混合信号中提取所需站址的信号.

扩频多址(spread spectrum multiple access) 简称 SSMA. 是采用扩展频谱技术来实现码分多址的一种码分多址类型.

通信卫星(communication satellite) 一种人造地球卫星. 即用于通信目的的一种人造地球卫星. 它起中继站的作用,即把来自地球站的信号经放大后转发给它所覆盖的地球站. 此外,通信卫星还可完成星上处理和星上交换等功能. 通信卫星按其稳定方式可分为自旋稳定卫星和三轴稳定卫星,稳定方式决定于卫星的结构. 按其运转轨道可分为赤道轨道、倾斜轨道和极轨道. 目前,大容量固定业务的实用通信卫星主要使用赤道轨道上的静止卫星,也称同步卫星. 它的理论轨道是和地球自转同步的轨道,卫星定位于赤道上空 35 786km 处,卫星运行周期为一个恒星日,即 23 小时 56 分钟 04 秒. 从地面上看去,它恰似停在空中静止不动. 而移动卫星通信所用的通信卫星多为倾斜轨道和极轨道.

对流层散射通信(tropospheric scatter communication) 无线电通信的一种方式. 即利用对流层对无线电波的散射作用进行的通信. 对流层指离地面 10～18km 高度以下的大气层. 在对流层中,不均匀的气象因素(如气温、温度等的变化)形成大小、形状各不同的旋涡状气团——湍流团. 湍流团内的气压、温度、介电常数与周围空气不同,当无线电波遇到湍流团时,会向各个方向散射,这样可在一端发射电磁波,而在远方一端用接收机接收散射回地球的电磁波,以完成通信. 对流层散射通信的特点是:不受电离层骚动、太阳黑子活动的影响,通信距离较远,可达 600km,通信容量大,可进行上百路通信;可靠性较高,工作频率为

$$100\text{MHz} \sim 10\text{GHz} \quad (1\text{G} = 1000\text{M}).$$

但对流层散射通信的传输损耗大,收发信设备复杂,天线庞大,成本较高. 它常用于无法建立微波中继站的地区,如跨海、沙漠、湖泊、高山等地区.

软件无线电(software radio) 一种新型无线电技术. 指利用软件技术实现无线接口上可重新配置的无线电技术. 20 世纪 90 年代初,美军为了解决多兵种间无线电通信设备的互通、互联和设备的统一,开始研制了一种多频段(2MHz 至 2GHz)、多模式、多标准、采用可编程软件实现的无线通信系统“易通话”. 它被认为是全世界最早实现的软件无线电雏形. 但是,目前研究最多的是用软件定义(SDR)的无线电设备,其主要特点是将所需的不同业务与技术要求都用软件定义功能实体(模块)来实现. 主要可以划分为三大部分:通用的宽频带射频模拟段,高速中频模数转换段,可编程的基带数字处理与配置单元. 其核心是通用化、数字化和软件可编程化. 将来,软件无线电可能是一种“虚拟”的无线电设备,即可通过一个超高速、宽带信息网即时下载所需软件,并将其动态组合成不同功能的无线电通信设备.

智能天线(smart antenna) 一类无线抗干扰技术. 是利用信号传输的空间特性抑制干扰,以空间入射角度来区分信号与干扰而实现空间滤波的一类抗干扰技术. 消除干扰可以从频域、时域和空间域不同角度来实现,智能天线侧重于从空间域来实现,它利用信号与干扰来自不同方向的特点,通过自适应控制天线阵列的方向图,使它在期望的信号方向上保持高增益,而在干扰方向上形成零陷,以达到增强信号、抑制干扰的目的. 20 世纪 60 年代,军事上首先利用这一原理成功研制了相控阵雷达,它可以自动形成多个无线波束来跟踪多个动态目标. 20 世纪 80 年代以来,人们又将它引入移动通信领域,并称它为智能天线. 它试图利用多个自适应波束跟踪每个移动用户. 目前实用的正式产品尚未达到可自适应波束跟踪的要求,而是采用预多波束自适应切换方式来实现. 智能天线技术主要包括天线阵列结构、由加权网络组成的多波束形成和由信号处理完成的各类自适应算法三大部分. 它是第三代移动通信的核心技术之一,也是今后空分的主要方向.

个人通信(personal communication) 一种通信手段. 它被认为是 21 世纪的主要通信手段. 虽然目前还没有统一、确切的定义,但是,从其基本要求来看,它被人们理解为能在任何时间、任何地点、可以同任何人建立各种主要通信业务的一种手段. 可见,个人通信应具有个人化、多媒体化、智能化的基本特色. 要实现它,首先,在全球网络层要形成统一制式的网络结构,至少可以互联、互通;其次,要实现全球至少每人一个地址号码,以满足全球几十亿人都可以通信的基本要求;第三,要保证个人通信的安全性,具有身份认证、通信保密的基本要求. 显然,这是一个在较长时间内的奋斗目标. 目前,国际电信联盟(ITU)正在制定和实施的第三代移动通信 IMT-

2000系统是走向这一目标的一种最佳过渡方案.它将目前的数字式蜂房移动通信、移动式卫星通信、无线电话以及无线寻呼逐步统一为IMT-2000系统,初步实现个人化、多媒体化、智能化,并能实现在全球任何地点、任何时间、任何终端用户间的通信.

移动因特网(mobile Internet)　一种狭义因特网.最大的、开放式的、并由众多网络互联而成的计算机网络称为因特网(Internet),又称大因特网或广义因特网.移动因特网则是利用移动通信手段来实现的狭义因特网.它是信息领域发展最快的两个领域(IP因特网与移动通信)相结合的产物,将给正在形成的信息社会产生巨大的影响.移动因特网诞生以来,短短几年已迈出了三大步:利用手机作为传输手段,再利用传统有线电视连结,比如红外和蓝牙技术连结笔记本电脑,实现全无线移动上网;利用专门无线应用协议,比如WAP,进行手机直接上网.目前实现移动因特网最成功的是日本I-mode体制.即将运营的第三代移动通信实现宽带、多媒体业务直接移动上网服务,具有广阔的应用前景.

电　视

电视(television)　用电的方法实时传输活动视觉图象的装置.即利用光电转换技术、电信号传播技术及电光转换技术,将活动视觉图象传送到远方,并即时重现图象的装置.电视通信系统的主要组成部分是电视摄像机、电视接收机和传输系统.电视摄像机是一种光电信号转换设备.它将需要传送的图象通过对每帧画面从上到下、从左到右、逐点、逐行、逐帧地扫描转化为电信号.携带图象信息的电信号经传输媒体送到电视接收机.电视接收机是一种电光转换设备.它将电信号依次逐点、逐行、逐帧同步地还原成光信息,在荧光屏上显示出整个图象画面.每帧画面分解成的行数和点数越多,重现后的图象就越清晰.中国电视广播标准采用每秒钟传送画面25帧,每帧画面625条线.

1884年,发明了圆盘旋转扫描器,用光电管把图象的序列光点转变为电信号,实现了最原始的电视图象传输和显示.1925年和1926年相继实现图象粗糙的机械扫描电视系统.1930年发明的电子扫描系统和美国无线电公司(RCA公司)的电子束显像管,使电视进入了现代阶段.1937年在英国、1939年在美国开始了黑白电视广播.到20世纪50年代初期,黑白电视广播逐步在各国普及.中国在1958年开始黑白电视广播.同时在20世纪50年代初期,人们根据红、绿、蓝三种基色光相加可得到不同彩色感觉的原理,开始彩色电视的研究.美国最先试播一种与黑白电视不兼容的顺序制彩色电视.到1953年采用了NTSC兼容制彩色电视制式,1954年正式广播.联邦德国、法国相继于1963年、1966年确定了兼容制的PAL与SECAM彩色电视制式,与美国的NTSC制式并列为世界三种彩色电视制式.它们分别在世界各国和各地区得到采用.大多数国家从20世纪60年代后期转向彩色电视广播.中国从20世纪70年代初开始发展彩色电视广播,采用PAL-D制式.

目前,由于大规模集成电路、微处理器和数字技术的应用,以及新型广播业务的出现,电视机的性能和结构发生了很大的变化.新型的彩色电视机增加了许多功能,例如,遥控随机选台;屏幕上随时显示频道号码、音量响度和时间;预编程序,在预定时间自动开机、关机、转换频道和播送节目;多画面显示;接收多伴音、立体声、文字广播、静止图象广播、传真数据传输等多种电视广播节目.随着电子技术的发展,电视接收机将向多功能化、微型化、数字化、智能化方面发展,以适应多方面的需要,如卫星直播电视接收机、高清晰度大屏幕电视、平板电视、电视录像组合机、电视游戏机以及立体显示等.电视接收机还将与视频数据检索系统和家用计算机结合,成为家用信息显示终端.目前,电视已广泛应用于广播、通信、工业、交通运输、医疗、教育、科学研究和军事等方面.

黑白电视(monochrome television)　一种以单色(黑白)画面传送和显示活动自然图象的电视.它是按照被摄景物的不同亮度而由呈现不同灰度的黑白(单色)像素构成画面的图象传输系统.黑白电视技术简单,在早期广播电视中以及工业电视中应用相当广泛.

彩色电视(colour television)　一种以彩色画面传送和显示活动自然图象的电视.在摄取景物时,通过分光系统将一帧彩色图象分解为红、绿、蓝(R,G,B)三幅基色画面,由摄像器件将其转化为三基色电信号,将它们编码成亮度信号与色度信号的组合后传送出去.彩色电视接收机收到信号后,经解码和电光转换,使三幅基色画面组合成一帧彩色图象,重现景物的亮度和色彩.各国传送彩色电视广播信号的方式(称制式)有所不同,目前主要有NTSC制、PAL制、SECAM制.这三种制式均能"兼容"各自的黑白电视:即其黑白电视接收机能接收其彩色电视广播,重现的是黑白图象;彩色电视接收机可接收黑白电视广播,重现的也是黑白图象.

广播电视(broadcasting television)　点对面通过传输介质向公众传输图象信号和声音信号的一种电视.广播电视信号用米波、分米波作为传输介质发送时,受到电波视距传播特性的限制.广播电视信号要实现远距离的传输,可以通过微波、光缆干线或卫

星传送，实现广播电视的大面积覆盖.

闭路电视(closed circuit television)　亦称为有线电视. 信号传输限制在非开路空间内的电视. 即通过射频电缆、光缆传送和分配图象、声音信号和数据信号的电视，其原理与广播电视相同. 这类电视在20世纪40年代首先应用于工业部门. 目前，已广泛应用于工业、科学研究、交通运输和医学等方面. 它的主要功能是实现监视、数据传输、信息记录、测量记录等. 此外，还可用于教育、娱乐等电视节目的传输.

电视制式(TV system)　电视广播的技术规范或标准. 目前世界上有100多个国家与地区使用各种不同的制式. 国际无线电咨询委员会(CCIR，现改变为ITU-R)把这些制式划分为13类基本制式. 彩色广播电视中彩色信号编码的制式主要有三种：NTSC制、PAL制和SECAM制. 中国现行的彩色电视广播制式是PAL-D制.

NTSC制(national television system committee system)　一种兼容其黑白广播电视的彩色电视制式. 即被传送的彩色图象在发送端被分解为红、绿、蓝(R,G,B)三个基色信号. 为了与黑白电视兼容，将它变换为一个亮度信号(即Y信号)与两个色差信号(即I,Q信号)作为传送的信号. 利用高频混合原理，将色差信号频带限制在较窄的频带内而亮度信号则用全频带传送. 两个色差信号对副载波正交平衡调幅后，利用频谱交错原理插入到亮度信号中一起传送. 这里副载波频率f_s选为半行频的奇数倍，例如在美国$f_s=3.579545\text{MHz}$. 此外，在传送信号的行消隐期间，插入一个色同步信号，作为接收端副载波频率和相位的基准. 这样，接收端用同步检波法就可分离出两个色差信号；连同Y信号，可以恢复成三个基色信号，它们的低频分量是原基色信号的低频分量，而高频分量是亮度信号的高频分量. 因此，在接收机上重现的是一幅大面积着色的彩色图象. 由于NTSC制中色度信号的相角代表被传送彩色的色调，而在接收端没有采取抗干扰措施，所以它对相位畸变较敏感，对传输设备的相位特性的要求也就比较严格. 但是，NTSC制接收机的解码电路比较简单. 1953年，美国将NTSC制定为国家彩色电视标准. NTSC是美国国家电视制式委员会(national television system committee)的首字母缩写. 1954年，正式用于广播电视. NTSC制也用于加拿大和日本等国.

PAL制(phase alternation line system)　一种改进的电视制式. 是为降低NTSC制的相位敏感性而发展的一种电视制式. 于1967年正式用于广播电视，也用于英国和中国等国. PAL是逐行倒相(phase alternation line)的首字母缩写. 这种制式沿用NTSC制的基本原理，所不同的有以下几点：

1. 采用U,V色差信号而不用I,Q信号，而且U,V信号均用1.3MHz带宽，残留边带方式传送.

2. 在发送端将副载波色度信号V逐行倒相(逐行反极性)，而在接收端再使其极性复原. 采用这一措施后，使得由相位误差引起的色调变化，在相邻行上具有互补性质. 因而用一根延迟线将两行信号平均后，可使色调不受影响，而只是色饱和度略有下降. 这是PAL制解决NTSC制相位敏感性的基本原理.

3. 为了尽可能减少亮度和色度信号之间的相互干扰，副载波频率f_s采用以下关系式：

$$f_s=\left(284-\frac{1}{4}\right)f_H+\frac{f_V}{2},$$

式中f_H为行频，f_V为场频. 例如当$f_H=15625\text{Hz}$，$f_V=50\text{Hz}$时，$f_s=4.43361875\text{MHz}$.

4. 彩色同步信号中除了副载波频率信息和恒定相位分量外，还有一个逐行倒相的分量，以便起到副载波定相和PAL识别的双重作用. PAL制接收机中的解码电路比NTSC制的复杂，造价有所增加.

SECAM制(sequential colour and memory system)　一种改进的电视制式. 即为改善NTSC制的相位敏感性而发展的一种兼容的彩色电视制式. 1967年，在法国正式用于广播电视，还用于俄罗斯和许多东欧国家. SECAM是顺序传送彩色和存储(sequential colour and memory)的首字母缩写. 它仍由三个基色信号合成亮度信号和两个色差信号. 亮度信号每行都传送，但两个色差信号D'_R和D'_B则逐行轮流传送. 这样，每一行都少传送一个色差信号. 由于相邻行上的信号大体相同，在接收机中采用一根64μs的延迟线，将上一行的色差信号延迟到这一行来使用，因此每一行仍然可以有一个亮度信号和两个色差信号，经矩阵变换后也可以重新获得三个基色信号. 由于每行只需传送一个色差信号，可以采用副载波调频方式，不必采用NTSC制的正交平衡调幅方式，因而在克服相位敏感性方面取得显著效果，而且受传输中的多径接收的影响较小，但同时也产生了在高饱和度彩色垂直过渡处出现12.5Hz闪烁现象和兼容性较差等缺点.

电视制式转换(TV system conversion)　不同电视制式间的转换. 有两种方式：一是采用光电转换方式，即先将一种制式的电视信号还原成光信号，然后再将光信号转换成另一种制式的电信号. 这种方式将带来图象质量的下降，且不适于不同制式间的往返转换. 另一种是数字方式，它将电视信号中的亮度信号和色差信号先解码，然后分别予以取样处理. 优点是图象转换的保真度高，且适合于不同制式间的往返转换. 缺点是电路复杂，设备费用较高.

彩色电视制式转换(colour TV system conversion)　不同彩色电视制式间的转换.即采用不同彩色电视制式的国家,为了相互交换电视节目,就必须将一种制式的电视信号转换成另一种制式的电视信号.对电视信号进行这种处理就称为彩色电视制式转换.场频、行频相同的彩色电视制式之间的转换,只需要改变彩色电视信号的编码方式.完成此种制式转换的设备称变码器.对于场频、行频不同的彩色电视制式之间的转换,除了上述变码过程外,还必须改变扫描的标准.这种制式转换设备目前都是数字式的.数字式场(行)存贮的转换是采用数码技术和场(行)内插技术来实现的.

变码器(transcoder)　见"彩色电视制式转换".

电视频道(television channel)　电视术语.广播电视系统中,指规定给各个广播电视通道使用的频率范围指配.每个电视频道所占用的频带宽度,各国规定有所不同,中国规定为 8MHz.中国规划的各电视频道的频率范围如下表:

频道序号	频率范围/MHz	频道序号	频率范围/MHz	频道序号	频率范围/MHz
1	48.5～56.5	17	502～510	33	670～678
2	56.5～64.5	18	510～518	34	678～686
3	64.5～72.5	19	518～526	35	686～694
4	76～84	20	526～534	36	694～702
5	84～92	21	534～542	37	702～710
6	167～175	22	542～550	38	710～718
7	175～183	23	550～558	39	718～726
8	183～191	24	558～566	40	726～734
9	191～199	25	606～614	41	734～742
10	199～207	26	614～622	42	742～750
11	207～215	27	622～630	43	750～758
12	215～223	28	630～638	44	758～766
13	470～478	29	638～646	45	766～774
14	478～486	30	646～654	46	774～782
15	486～494	31	654～662	47	782～790
16	494～502	32	662～670	48	790～798

电视接收机(television receiver)　简称电视机.广播电视系统的一种终端设备.即把天线接收到的高频电视信号还原为视频图象信号和声音信号,分别加给显示器件和扬声器,重现图象和重放声音.

黑白电视接收机(monochrome television receiver)　简称黑白电视机.广播电视系统的一种终端设备.即用来接收电视台播送的黑白电视节目,也可以接收兼容制式的彩色电视节目,但只能重现黑白图象及其声音.它由接收天线、高频调谐器、图象中频公共通道、声音通道、扫描系统和电源部分等组成.

彩色电视接收机(colour television receiver)　简称彩色电视机.广播电视系统的一种终端设备.即在接收电视台播送的彩色电视节目时能重现彩色图象及其声音,当接收播送的黑白节目时能重现黑白图象及其声音的接收设备.

彩色电视机的电路由通道、彩色解码器和显示器三部分组成.通道部分包括高频调谐器、图象中频公共通道和声音通道.

电视波段(TV band)　供广播电视使用的波段.由若干个电视频道组成电视波段.甚高频段(VHF)分 48.5～92MHz 和 167～223MHz 两个频率范围,特高频段(UHF)为 470～958MHz 频率范围.

测试卡(test card)　电视术语.指出现在电视屏幕上的图案.是一种用以测试(主观评定)电视图象质量的画面图案.

常用的彩色电视测试图案四周是黑白矩形块组成的宽度比为 4∶3 的边框,用以检查调整行(水平)和场(垂直)的扫描幅度以及图象中心位置和同步分离是否正常.边框内大小相同的灰底白线正方格能反映出扫描线性、幅度大小和几何失真.对于彩色电视机,白线条清晰表明会聚良好,有灰色背景表明色纯正常.中央大圆呈正圆形表明屏幕中部的行、场扫描线性良好,大圆边缘光滑表明隔行扫描正常.圆内电视台标记下面从左到右是男、女肤色带条.再下是五组水平清晰度线组,从左到右依次为 140,220,300,380 和 450 线,用于调整频率微调和聚焦,使之达到最清晰的程度.清晰度线下是从黑到白 6 个灰度等级,用于调整对比度和亮度,使各级灰度的层次差别明显.图的中部是黑背景上的白十字线,用来确定图形的中心位置、判断静会聚是否正常和有无并行,微调场频可使隔行扫描良好.中部靠下是彩条,调节色饱和度和色调可使各颜色鲜艳、保真.彩条下边是黑背景上的白矩形,用来检查接收机的频率特性.两侧的黑色竖线可检查有无重影.

电视摄像(image pick-up)　电视术语.指摄取景象将其变成电视信号的过程.电视系统摄取景物的景象,并把它转变为电信号的光电转换过程称为电视摄像.通常是用电视摄像机产生彩色电视图象信号.

电视摄像机(television camera)　广播电视的重要设备之一.即把光学图象转换成符合规范的视频电信号的设备.一般分为演播室、便携式、无线式和特殊环境专用等类型电视摄像机.摄像机头包含光学系统、摄像器件、扫描电路及预放器等,控制单

元包含消隐、同步、视频处理和控制电路，供操作者调整亮度、色度和某些校正以使图象最佳．当前的电视摄像机已全部固体化，还具有部分或全部的自动调整功能，体积、质量都大大减轻，目前已开发出全数字化彩色电视摄像机，图象质量更高．

电视录像(video recording)　电视术语．指记录电视图象和声音信号的方法．主要方式有磁带录像和光盘(VCD)录像，目前已生产出可擦录的视频光盘，将逐渐取代磁带录像方式．

磁带录像机(video tape recorder)　简称 VTR．广播电视的重要设备之一．是利用磁性物质(如磁带)记录电视图象和声音信号的设备．广泛应用于工业、医学、教育、军事、科研等方面．

数字电视(digital television)　用数字信号形式表示图象信号，并采用数字技术对视频信号进行加工、记录、编辑、存储、调制和传输，以及接收端的解调和处理而显示出彩色图象的电视．如果从摄像播送端到接收显示端的整个系统中均采用数字信号处理技术处理视频信号，便构成了全数字广播电视系统．如果在摄像播送端的电视中心台采用了数字信号处理技术，而向观众传输时先将数字信号变换回模拟信号，然后依然按模拟的 PAL 制方式实施编码和调制，进行传输，接收端接收的仍是 PAL 制模拟视频信号，则只是采用了数字演播室系统．

对于全数字广播电视系统，相应地需要由数字电视接收机来接收已数字调制的高频电视信号，经变频、放大、检波等处理后可提供出质量十分优良的电视图象．数字电视系统的优点是其图象质量高，原则上，当信号传输条件优良时，观众能看到电视台播的图象质量；数字电视机的功能大为强化，能满足用户的多种要求，例如收看加密的付费电视，点播电视节目，以及给出观众需要的图文数据和网上信息等(如影视节目、交通信息、行情等)．

在实现数字广播电视中，技术方面的处理步骤是：先使数字摄像机中摄像器件最初给出的模拟视频信号通过模拟数字变换器变换成数字视频信号，而后完全在数字信号领域内进行各种信号处理，一直到视频信号加到显像器件之前再通过数字模拟变换器变换回模拟视频信号，供显像器件显示出彩色图象来．伴音部分也是如此．

所以，全数字广播电视系统中，所有各环节上的设备都应是数字式的，诸如是数字摄像机、数字录像机、数字特技机、数字非线性编辑装置、数字切换台、视频服务器、数字调制器和数字电视机等，它们与模拟广播电视系统中的设备是不相同和不兼容的．因此，数字电视系统是新一代的广播电视．

将现有的模拟广播电视系统升级成数字广播电视系统进行广播，特定地称为标准清晰度电视．以中国的模拟电视为准，数字化后的标准清晰度电视，一帧图象的总像素为 720(水平)×576(垂直)＝414720，即约 40 万个像素，而扫描参数是与原来的 PAL 制一样的．

高清晰度电视(high definition television)　简称 HDTV．一种数字电视．景象细节在显示器件上重现出的清晰程度远高于现有的常规电视图象清晰度(也称为标准清晰度)的电视．高清晰度电视是由于现有的模拟制式广播电视其图象质量不够高，不能满足人们所期望的观看接近电影质量的画面而研制成功的．从 20 世纪 70 年代提出设想，到 20 世纪 90 年代出现可具体实施的 HDTV 广播电视标准，其间主要是日本、欧洲和美国投入了力量制定 HDTV 系统的优化参数，20 世纪 90 年代以来，中国也积极参与了数字高清晰度电视的研发．目前，规定的一帧图象的像素数为 1920(水平)×1080(垂直)＝2073600，即约 200 万个像素，它大约是 SDTV(标准清晰度电视，即数字化的常规电视)一帧图象像素数的 5 倍；规定的图象帧型比(宽高比)从现有电视的 4：3 增宽至 16：9．所以，HDTV 的图象是宽屏幕、高清晰的．对于 HDTV 图象，要求接收端用大尺寸的显示器件(对角线大于 80cm)进行显示，适合于观众近距离(图象高度的 3 倍)观看．此时，观众对 HDTV 图象内容会具有身临其境的现场感，极富吸引力．这正是人们开发 HDTV 所追求的目标．

由于 HDTV 图象的像素数多，相应地传输中所需的视频信号带宽将比现有电视的高出许多倍．为此，必须在数字信号领域内实现数据率(或称码率、比特率)的深度压缩，使得在一个现有模拟电视的频道带宽(例如 6，7 或 8MHz)内能够传输一路高清晰度电视信号．这只能在数字化的基础上才能做到．所以，HDTV 必定是在数字电视技术上实施的．HDTV 与 SDTV 相比较，在数字信号的处理技术上并无本质差异，但在电视图象质量上人们会感觉高出一大等级．另外，在 HDTV 中，规定了声音信号采用 5＋1 声道环绕声的标准，使得在图象呈现上观众有临场感的同时，在声音呈现上也有相应地逼真的临场感．5＋1 声道是指观众的前左、中、前右、后左和后右方各布置一路信号带宽为 20kHz 的扬声器，另外在前方有一路信号带宽为 160Hz 的低音扬声器．

接收高清晰度广播电视信号，需要有对应的高清晰度数字电视机．原理上，标准清晰度数字电视机或者 PAL 制式电视机不能够直接接收高清晰度数字电视信号，直接欣赏高质量的图象和声音．但实践中，用户家中已有的 SDTV 或者 PAL 制式电视机可以依靠配置附加的机顶盒，使接收到的高清晰度电视信号变换成 SDTV 信号或者 PAL 信号后再传

输给已有的电视机，从而收视到降低些质量的图象和声音.

标准清晰度电视(standard definition television)　简称SDTV. 一种全数字电视. 即取代常规模拟电视广播系统(包括NTSC制、PAL制和SECAM制)中的复合信号编码方式，按照国际规定的电视演播室数字分量编码标准，使它们的亮度信号Y、红色差信号C_R和蓝色差信号C_B分别进行模拟-数字变换而构成的全数字电视广播系统(参见"数字电视")，特定地称为标准清晰度电视(SDTV). 在SDTV中，保留了原来模拟系统中的扫描参数及格式(625行/50场/2：1隔行或525行/59.94场/2：1隔行)，对于数字分量信号在实施码率压缩后按时分复用方式将Y,C_R,C_B和数字声音组成的码流调制成高频信号向外传输，接收端使用数字电视机进行接收而恢复出彩色图象和高质量声音. 由于没有模拟系统中亮度和色差信号的复合编码处理，SDTV可以清除复合编码信号所带来的模拟系统中图象质量方面多种固有的原理性缺陷，能够使图象质量水平大为提高. 而且，通过码率压缩处理，可使一路SDTV信号在调制传输中占用的频带不超过2MHz带宽. 换言之，现有常规电视指配的一个频道内(6,7或8MHz)，容许传输4至6套电视节目，极大地提高了现有电视频道的频谱利用率，而图象质量并不比常规电视的差.

SDTV中的扫描参数和格式保持与模拟电视相同，国际标准规定其一帧图象的像素数为720(水平)×576(垂直)(对于625/50格式)或是720(水平)×480(垂直)(对于525/59.94格式)，所以，SDTV的图象清晰度并不会比常规的模拟电视高多少，因而属于标准清晰度水平.

接收端接收SDTV的电视广播信号时，使用符合相应标准的数字电视机. 当观众欲利用原有的PAL制式电视机进行接收时，需要配置机顶盒将SDTV信号解调并再编码成PAL制式信号或者亮度信号Y与色度信号C后再馈送给PAL制式电视机.

经模拟传输，如果只是在PAL制式电视机中对解调得到的复合视频信号实施模拟-数字变换，然后在数字信号领域内进行改善图象质量的处理，在一定程度上克服复合视频信号固有的一些缺陷，有助于改善电视图象质量. 然而图象质量依然不能与SDTV的相比. 此类PAL制式电视机属于数字处理电视机，而不属于真正的、直接接收数字调制信号的数字电视机.

直播卫星(direct broadcasting satellite)　简称DBS. 一种数字电视广播装置. 即利用地球同步卫星(静止卫星)转发器将地面上行发射台(上行站)定向发射的高频电视信号转发向地面以供地面接收站个体用户直接收看电视广播节目的技术，它可实施电视广播、声音广播和数据广播等. 它可使用11.7～12.2GHz内500MHz的Ku频段.

由于现在的卫星广播下行频率在12GHz的Ku频段上，且卫星转发器的发射功率为一二百瓦，地面EIRP(等效全向辐射功率)可达到50W左右，所以个体接收的碟形天线直径只需40～60cm，尺寸较小，安放方便.

一个接收天线同一时间可以对准天空中某一颗直播卫星，在一颗直播卫星上可安装24个转发器，每个转发器分配27MHz的带宽，在保证数据信号误码很小的一些编码参数下，每个转发器能传输大约20Mb/s的码率. 当一套电视节目压缩到大约3Mb/s码率(有良好的图象质量)时，一个转发器能够传送7套电视节目. 因此，接收天线对准该颗卫星时容许用户从150套以上的电视节目中选择收看所需的节目. 所以，直播卫星能够向各个用户直接提供数量多、质量好的电视节目. 由于地面上行站向卫星传送的是数字调制高频电视信号，因此，直播卫星系统属于数字电视广播.

有线电视(cable television)　简称CATV. 利用有线传输构成的电视系统. 有线电视是在20世纪中期首先应用于工业部门的闭路电视以及后来集团内部采用的共用天线电视系统基础上发展起来的，现在已通过遍布城市间和城市内由光缆和电缆组成的主干网和分配网，使有线电视网成为广播电视信号极重要的传输及覆盖手段之一. 同时，有线电视网向着宽带化逐步发展，从早期的300MHz、450MHz过渡到550MHz、750MHz的带宽，使得线缆内能够容纳更多的电视频道(每个频道带宽像地面电视广播中的一样，是8MHz). 300MHz系统可以安排27个频道，450MHz系统可以安排46个频道，而550MHz系统可以安排58个频道. 550MHz以上的频带预定分配给数字电视信号有线传输使用，一个8MHz频道内可传输4至6套码率压缩的电视节目. 因此，750MHz的有线电视系统能够向用户提供几百套电视节目，是开路的地面电视广播不能比拟的.

有线电视系统的一大特点是易于构成双向传输系统，例如采用频率分割方式，规定5～65MHz频段传送由用户向前端发出的上行信号，87～750MHz频段传输由前端向用户馈送的下行信号，65～87MHz为保护频带. 借此，通过有线电视机顶盒可实施交互式电视业务，拓宽有线电视的应用范围，诸如向用户提供NVOD(准视频点播)、PPV(按次付费节目)及文体信息和远程教育等多媒体数据服务. 有线电视网又可成为宽带网信息高速公路物

理传输媒体的一部分.

网络广播(webcasting)　以互联网作为媒体，在上面实施具有交互能力的视音频广播的一种信息传输方式. 它既有传统电视广播实时性和生动性的特性，又有互联网上可按需获取信息、联系起电视台与观众之间交互响应作用的特性. 实践上有两种技术，即下载播放技术和流媒体播放技术. 下载播放属于非实时传输方式，电视台先将连续的视音频节目预先采集，压缩成数据文件，存放于网络服务器的磁盘上. 响应于用户的下载请求，由服务器通过互联网将文件下载到用户的计算机磁盘上，下载完成后计算机就可以播放，并可以反复播放. 这种方式由于需要先下载存储文件，所以不能实现直播的要求. 流媒体播放将视频流和音频流以恒定的数据率传输，用户端采用边下载边播放的方式，实时地重现播送端的图象和声音，播放完的数据不保留于用户计算机磁盘上. 网络广播，一般特指流媒体播放.

流媒体播放形式一种是直播，另一种是点播. 直播是电视台实时播出的节目，同时也在网上传输，它时效性强，生动实际，用户可在第一时间获知信息，但不能只选看所关心的部分，而又一直占用着带宽资源. 点播是电视台将节目根据内容做成一个个片段，用户可以根据标题或分类选择感兴趣的片段，即时点击接收，任意收看收听. 点播更能适应用户的要求，较好地利用带宽资源，是流媒体播放技术中十分适合于应用的一种网络广播方式.

网络广播的带宽是制约可提供的图象和声音之质量及用户数量的瓶颈，也是减少流媒体传输中出现断续现象的关键所在. 对此，随着网络基础设施技术水平的提高，宽带接入网络成为保证网络广播道路畅通的必然手段，所以，宽带流媒体服务是广播电视系统在网络时代的一大发展趋势.

数据压缩(data compression)　一种数据处理技术. 即借助数据压缩技术，结合视觉特点，尽量去掉原始数据中的信息冗余以减少数据量、降低数据率的技术. 经数据压缩后，数字视频信号的数据率降低，因而存储容量要求减小，传输带宽可以减低(故数据压缩又泛称频带压缩)，使得数字电视广播系统的构成变得现实而可行.

国际上建立了多种适应于不同要求和适用于不同场合的数据压缩标准，有 MPEG-1，MPEG-2，MPEG-4 和 MPEG-7 等，其中 MPEG-2 应用于标准清晰度电视(SDTV)和高清晰度电视(HDTV)广播中，MPGE-4 应用于多媒体技术中. SDTV 和 HDTV 中应用 MPEG-2 数据压缩标准时能使压缩比在 5∶1 到 100∶1 的范围内按需调节，具体视压缩后的图象质量要求而定. 实际采用的压缩手段包括预测编码、变换编码和熵编码等运算方法. 压缩的依据是利用了运动图象帧内和帧间诸像素之间的空间相关性、时间相关性以及图象结构冗余和信息熵冗余等原理.

数据压缩技术一直有多方面发展，以追求码率高效而运算方便的压缩算法，除了在 MPEG-2 中应用的 DCT(离散余弦变换)编码方法外，还有小波变换编码、算术编码、子带编码、矢量量化编码和 K-L 变换编码等压缩编码方法.

非线性编辑(nonlinear editing)　简称 NLE. 以计算机硬件系统作为视音频数据的存储媒体，利用操作系统和用户图形界面，在多媒体数据编辑应用软件的支持下，对视音频素材进行不受其内容、时间顺序限制的各种加工、修改和处理的编辑方式. 非线性编辑在 20 世纪 90 年代初期进入电视台后期制作机房，它不同于传统的磁带录像编辑中线性时间方式(即节目内容的相对时间关系不能变)的插入编辑和组合编辑，非线性编辑是一种综合性技术，涉及电视技术及计算机技术的主要领域，包括视音频处理技术、数据压缩技术、数字存储技术、计算机图形学和网络技术等. 它不仅能够编辑视音频节目，还可以处理文字、图形和动画等多种形式的素材.

数字视频特技(digital video effect)　采用数字技术将输入的视频信号在电视屏幕的二维或三维空间中进行各种方式处理的技术. 它能使画面上诸像素组成的图象发生缩放、变形、分裂、旋转、多画面、油画效果、波浪效果、镜像效果、负像效果或马赛克效果等不固定枚举的二维特技形式，以及发生翻页、卷绕、球面、风飘等难以想象的三维特技形式. 这能够极大地丰富画面的表现能力，激发观众的兴趣. 数字视频特技的原理主要是通过对写入图象帧存储器的像素地址和从帧存储器中读出像素的地址做出一定的改变位置的运算来完成图象的缩放、移位和分裂等，或是对像素数据流按一定规律做出数值改变来实现多种二维效果，以及应用透视运算关系来做出多种多样的三维特技效果.

数字视频特技不同于计算机动画创作，计算机动画创作是按照构图设想来创造活动画面，而数字视频特技是对原有的视频画面根据创意进行特技效果的变幻处理，对原有画面做出艺术上的渲染修饰. 随着数字视频特技设备硬件的不断完善和软件的不断开发，也能使三维特技与动画系统结合在一起，将真实的运动图象与创作的动画景物融合起来. 数字视频特技机的发展延伸至今，又将数字切换台的功能包含进来，做成数字切换台与数字视频特技机相结合的产品，成为数字演播室节目制作中必须配备的装置.

虚拟演播室(virtual studio)　由现场动态、实时图象与计算机构成的三维虚拟背景图象构成的演

播室.电视节目的前期摄录由摄像机对准"空"演播室内蓝背景前演员的表演来进行,而所有道具和布景都由计算机内存储的预先制作的背景图象提供,使计算机产生的背景图象与摄像机拍摄的人物实时地合成,而且随着人物的走动或者摄像机的移动或镜头的变焦,前景的画面变化时背景图象能协调地改变画面结构,整个画面的构成如同人物在实际背景中活动一样,这便是虚拟演播室技术.虚拟演播室系统中,使现场实时图象与计算机三维图象的虚拟背景通过蓝色键电路完美地融合在一起,能有效地增强电视节目创作的自由度,也能一定程度地降低节目制作成本.

虚拟演播室的关键技术是:计算机的强大、快速计算能力和三维图象处理能力,以及摄像机操作参数(俯仰摇移和推拉变焦)的跟踪技术及传感技术.当摄像机操作参数改变时,必须高精度地感知其微小变化量,将它们传送给计算机,以便计算机能同步地计算出背景图象,且必须相应地跟踪变化的形态.传感跟踪的方式有两种,一种是传感器定位方式,另一种是图形识别跟踪方式.前者要求摄像机上安装多个运动传感器,分别及时给出各个操作参数变化量,并传输给计算机;后者要求在蓝背景上画出有一定规律的浅蓝色几何图案,计算机通过分析被摄图象"空"背景上几何图案的变形,可得知摄像机与演员间相对位置关系变化的信息,这时摄像机上不必安装传感器.两种方式各有优缺点,都已经实际使用.

视频信号(video signal)　一种电信号.是频率范围对应于可视图象信息,其内容可在显示器件上呈现出来供人们观看的电信号.广义地,信息内容可以是图象、图形、图案、动画和字符数据等,但广播电视中,典型地可由图象作为代表.因此,视频信号通常也称为图象信号.具体的频率范围:最低频率 f_{min} 低达 0Hz,即直流成分,它对应于静止图象的平均亮度大小;最高频率 f_{max} 视不同的视频系统而定,中国 PAL 制电视广播的国家标准中,规定 f_{max} 为 6MHz,而高清晰度电视节目制作中,三基色红、绿、蓝信号,如亮度信号的模拟信号标称 f_{max} 为 30MHz.f_{max} 越高,也即视频信号带宽 $\Delta f(\Delta f=f_{max}-f_{min})$ 越宽,呈现的图象能有更高的清晰度,愈鲜明地显示出图象上精微的细节.

广播电视系统中,视频信号一般指由景物图象原始形成的模拟电信号,它未经模拟-数字转换或高频调制处理,所以可特称为基带视频信号.相应地,经过模拟-数字转换后,常称之为视频数据(流),经过高频调制后,便成为视频(数据)已调制信号,它可通过有线、无线或卫星等媒体进行传输.PAL 制中,在 6MHz 基带的亮度信号 Y 频带内,通过 PAL 编码,将 1.3MHz 基带的红色差信号 $R-Y$ 和蓝色差信号 $B-Y$ 的副载波已调制信号(频率为(4.43±1.3)MHz,统称为色度信号)加入至 Y 频带的高端后,所形成的信号称为复合视频信号或彩色全电视信号,以表明是亮度(Y)基带视频信号与色度(C)信号两者的合成信号.

视频服务器(video server)　电视台内视频网络的重要组成部分.是大存储容量、高输出带宽的视频数据存储装置.视频服务器所存储的大量节目素材便于各个部门共享,供各处的人员进行监看、编辑和加工等处理.完工的成品节目放在视频服务器内,可应用于实时播出或交互式的视频点播(VOD)中.概括地说,视频服务器能向用户提供视频数据存储及读出的服务,它响应于用户的命令,作为信息的存储和供给载体,可满足用户对视频数据高速度、多通道的随机即时访问.视频服务器不主动地与用户通信会话,但它可为用户提供各样的服务类型,诸如非线性编辑、电视节目、广告串联播出、浏览检索及视频点播等.

不同的应用场合,要求视频服务器有不同的特性,包括容量、带宽、可靠性、通道数目、存取速度、构成和物理媒体等.最常见的物理媒体形式是硬盘阵列视频服务器,采用 RAID(独立磁盘冗余阵列)结构,能够有高性能的主处理器、高速度的数据总线、多通道的输入/输出端口、合适的存储容量、可靠的数据存取和完善的操作控制系统等.在非线性编辑之类的应用中,图象质量要求高,需采用低压缩比的数据压缩技术,其视频服务器的存储容量不必太大;而实时播出之类的应用中,在保证一定的图象质量下,可以采用高压缩比的压缩技术,其视频服务器的存储容量需相当大,要求提供多通道以满足长时间的连续节目播出.

数字压缩JPEG (digital compression JPEG)　一种通用国际标准.JPEG 是 Joint Photographic Expert Group 的首字母缩写,即联合图片专家组,它是国际上 ISO/ITU-T 联合技术委员会制定的关于连续影调(灰度和彩色)静止图象数字压缩的通用国际标准.国际标准文件编号为 ISOCD10918.其应用领域包括计算机图画制作、彩色摄影、彩色传真和彩色印刷等静止图象的存储和传输.显然,数字压缩 JPEG 标准是针对静止彩色图象进行数字压缩的,所以它应用的基本压缩处理方法是帧内冗余信息的压缩,即画面内的二维空间信息压缩,而未利用运动图象相邻帧之间的帧间冗余信息压缩,即没有降低前后帧画面内的时间相关性冗余.然而,这正是数字压缩 JPEG 标准的适用对象所决定的.如果将它应用于电视图象上进行数字压缩,则压缩算法仅限于帧内数据信息,压缩比大小可根据图象质量要求予

以恰当调整，随不同的应用场合能够在2：1至50：1的范围内灵活地选取．电视演播室常用设备的非线性编辑系统中，为了得到帧精度的编辑效果，普遍采用了JPEG压缩技术，但压缩速度需满足25帧/秒的电视帧频值（数字压缩JPEG标准原本没有关于压缩速度的要求），因此，这种压缩技术专门称之为M-JPEG（Motion-JPEG）．

数字压缩JPEG中规定了两类压缩算法，一类以DCT（离散余弦变换）为主体，另一类以差分脉码调制（DPCM）为主体．非线性编辑系统中采用了第一种压缩算法，得到压缩比变化范围大、质量水平高低不同的图象．在DCT算法中有三个基本步骤：

1．对8×8的像块进行DCT变换，将空间像素值变换成空间频率系数值；

2．根据人眼视觉特性，对64个空间频率系数（DCT系数）进行加权量化；

3．对量化后的DCT系数，在“元”字形读出后，进行熵编码．

通过改变加权量化时的量化步长，可以调整压缩比．压缩比为2：1时，解码后的图象与原始图象相比，并无质量差别；6：1至8：1时，能满足良好图象质量的要求；30：1至50：1时，适用于脱机编辑场合．数字压缩JPEG的解码算法是与编码相对称的逆处理过程，需要有同等的数字处理能力和处理时间．这一点与MPEG（运动图象专家组）标准的不对称压缩方式（解码算法比编码算法简单得多）有重大差别．

有线通信

有线通信（wire communication）　一种通信方式．即利用电信号或光信号在受限空间的电导线或光导线中传送信息的通信方式．其中，利用电信号在导电金属线中传送信息的通信方式，称为有线电通信，例如架空明线、对称电缆、同轴电缆等．利用光信号在光导纤维中传送信息的通信方式，称为光纤通信．在20世纪30年代，建立了第一条电报线路．在20世纪40年代，建立了莫尔斯电报系统．在20世纪60年代，第一条横跨大西洋的海底电缆铺设成功．在19世纪70年代，发明了电话，之后迅速推广应用．在19世纪90年代末，各国建立起人工交换的有线电话通信网．在20世纪20年代，有线载波通信得到很大发展，并出现了自动电话交换．在20世纪60年代，数字通信开始发展，迅速形成取代模拟通信之势．在20世纪70年代，出现了数字程控交换．最初的电报电话通信线路，采用架空单根铁线．1877年开始采用硬铜线，并用双根铜线构成电话回线．不久，研制出铜芯线铅皮电缆．在20世纪30年代，研制了同轴传输线电缆（同轴电缆）．1960年发明了红宝石激光器．1970年成功研制低损耗光纤．之后，大量生产了以光纤为传输线的光缆，从而使光纤通信飞速发展起来．

人工交换（manual switching）　一种由人工操作来完成话路接通和拆线的交换方式．人工交换按其电源供给方式的不同，可分为磁石制交换和共电制交换两种．前者，通话和信号电源分别由用户电话机和交换机各自供给，后者通话和信号电源由交换机统一供给．通常人工交换设备由用户电路、塞绳电路、座席组成．

自动交换（automatic exchanging）　一种完全由机械或电子设备来完成话路接通和拆线的交换方式．自动交换分为电路交换和存储转发交换两种．自动电话交换的接续过程是由用户在电话机上拨号或揿动按钮来控制交换动作．其接续速度比人工交换显著提高．自动交换设备经历了步进制、机动制、纵横制和电子制几个发展阶段．自动电话交换机从信息传递方式上可以分为：模拟交换机、数字交换机、布线逻辑控制交换机、存储程序控制交换机等．

布线交换（wired logic control switching）　一种交换技术．即采用布线逻辑控制的交换方式．其控制设备是由一些具有不同功能的控制电路组成，各个控制电路通过进行一定的布线连结而具有一定的逻辑操作功能，在外来信号的控制下，完成交换．

ISDN交换（integrated services digital network switching）　实现ISDN用户接续与拆接的交换方式．ISDN支持的交换功能包括64kb/s电路交换功能、大于64kb/s电路交换功能、分组交换功能和无交换功能，其中64kb/s电路交换功能是ISDN的基本功能．ISDN交换由ISDN交换机实现，ISDN交换机是在现有程控数字交换机上改进而成的，新开发的功能有：

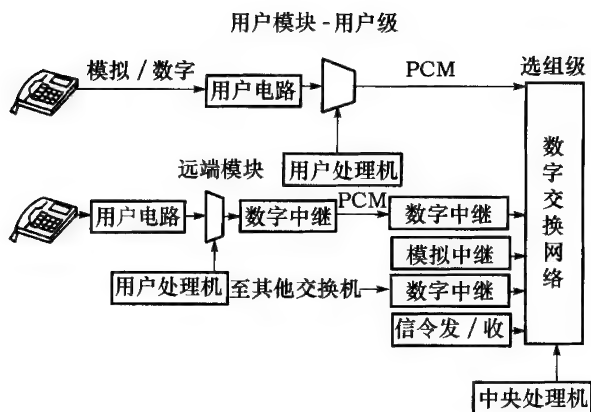
1．新的数字用户-网络接口，其中包括基本接口和一次群速率接口．基本接口上有2个B信道（64kb/s）和1个D信道（16kb/s），一次群速率接口上有30个B信道（64kb/s）和1个D信道（64kb/s）．其中B信道用于传送用户信息，D信道主要用于传送控制信令．

2．D信道处理功能，即实现D信道协议．

3．7号信令系统中的ISUP（ISDN用户部分）．

4．分组交换处理功能，或提供到分组交换网的接口，或在ISDN交换机中增加分组处理模块来完成此功能．

交换机（exchange）　通信网的重要设备之一．它负责完成通话用户之间的连结，并在通话终了时拆线．最早出现的交换机是人工交换机，由话务员控制交换接续，以后由机电设备取代人工，出现自动交



换机. 目前广泛使用的数字程控交换机用计算机控制电话交换的接线过程，利用数字技术将话音转换成数字信号进行传输和交换. 具有 ISDN 功能的数字程控交换机（又称 ISDN 交换机）能提供包括话音及非话音业务在内的综合业务交换. 数字程控交换机主要由硬件系统和软件系统两部分组成，硬件系统构成如图所示；软件系统则包括运行软件和支援软件两部分，前者实现对交换机的控制、管理和维护等功能，后者为开发运行软件提供各种支持功能.

程控交换机（stored-program control exchange） 一种自动交换机. 它是由电子计算机按照预先编好的程序来控制接续的自动交换机. 程控交换机又可分为：空分式程控交换机（采用空间分割交换方式）；时分数字式程控交换机（采用时分数字交换方式）. 程控交换机的控制方式灵活机动，如要更改或增加交换机功能，只需改变或增加相应的程序，而不必改动交换机的布线和硬件设备.

电路交换（circuit switching） 一种信息交换的方式. 电路交换为通信双方的用户提供电路连结. 即当用户之间要通信时，通信网中的交换机立即在用户之间建立一条电路，供用户互通信息，通信结束时，交换机将电路拆除.

存储转发交换（store and forward switching） 一种信息交换的方式. 即先把发端用户的信息存储在交换机中作适当的处理，一旦输出口有空，再把它转发给收端. 这种方式主要用于公众电报和数据交换. 存储转发交换又可分为电文交换和分组交换.

帧中继（frame relay） 简称 FR. 一种快速分组交换网络. 它是在分组交换数据网基础上发展起来的，并采用异步时分复用和存储转发交换方式. 为了减轻网络节点机的处理负担，帧中继简化了分组交换协议，节点机除了纠错码检验外，不参与任何差错控制和流量控制工作，因此节点机的处理速度大大提高. 目前，帧中继传输速率可达到 E_3（34Mb/s），节点时延降低到 2ms 以下. 帧中继的协议由 ITU-T 制定，如图 1 所示.

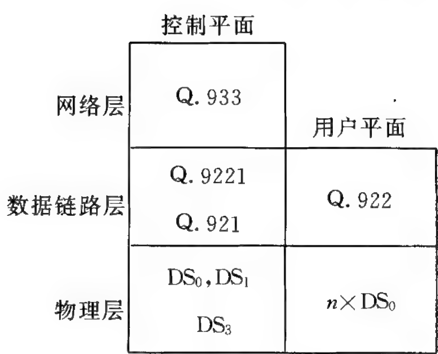


图1 FR协议

帧中继网的数据链路层协议采用 ISDN 网 D 信道链路控制协议 LAPD，帧结构由 ITU-TQ. 922A

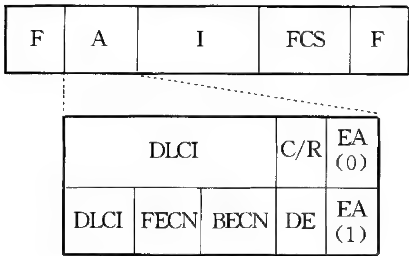


图2 FR的帧结构

建议规定，如图 2 所示. 帧结构由 5 个字段组成：F（标志段）作为帧的同步与定界在帧的开始和结尾处使用，A（地址段）主要用来表示数据链路连结的标志，I（信息段）是用户数据，FCS（帧校验序列）是纠错码，用于发现和纠正一帧中的传输错误. 地址段的详细结构表示在图 2 的下半部分，其 DLCI 是数据链路连结标志. 帧中继源端与目的端之间的连结，可能经过多个中间节点，因而包括多段 DLCI 链接，节点间最多可有 1024 条数据链路. C/R 是命令/响应比特，用来表示该帧是命令帧还是响应帧. FECN 及 BECN 分别表示前向及后向拥塞警告通知，用户终端或节点机可根据信道拥塞程度，自动调节发送数据量，实现流量控制. EA 为地址扩展位，DE 是帧舍弃许可指示，当发生拥塞时，可舍弃 DE＝1 的低优先级帧. 帧的总长度是可变的. 网络依据地址信息，寻找优选路由并送往目的端. 当帧中继网中某节点业务量过大临近拥塞（超过一个门限值）时，由该节点分别向目的端及源端发出 FECN 及 BECN，通知源端限制业务信息流. 如果拥塞状况继续恶化，则舍弃 DE＝1 的帧.

目前，帧中继网主要提供永久虚电路（PVC）服务，并已开发出交换型虚电路（SVC）. 帧中继典型应用是局域网间高速互联，局域网通过路由器（R）与帧中继网连结，互联速率可采用 $n\times 64$kb/s 及 E_1（2. 048Mb/s），甚至可达 E_3（34Mb/s），从而实现局域网-帧中继-局域网（LAN-FR-LAN）高速互联

结构.帧中继网也可用于图象或图文传输.此外,可利用帧中继网作为分组交换网的中继网,将帧中继高吞吐量及低时延性能与分组交换网高可靠性及差错控制能力相结合,提供高效可靠的传输与交换.

异步传递方式(asynchronous transfer mode) 简称ATM.宽带综合业务数字网(B-ISDN)的传递方式.它综合了传输、复用及交换技术,成为通信和计算机网络运营和制造商关注的热点.许多网络公司及电器设备厂家已推出ATM交换机及附属设备,国际标准化组织也制定了一系列相关的标准.ITU-T I.121建议了宽带综合业务数字网总体框架,并确定ATM作为B-ISDN的核心技术,并提出由现有网向ATM网过渡的策略.ATM网络具有高带宽、低延迟、低时延抖动、保证服务质量等优点.ATM网的接入速率可选择155.52Mb/s或622.08Mb/s,即B-ISDN网T参考点速率,能够适应所需宽带业务的需求.ATM网采用多种先进的业务流量控制及拥塞控制技术,如业务量整形、连结接纳控制(CAC)、用户参数及网络参数控制(UPC/NPC)、优先级控制等.

异步传递方式采用异步时分复用,并以信元作为传输与交换的基本单元,信元长度为53个字节,包括5个字节的头部和48个字节的用户信息段(又称净荷).这种短小的定长信元,便于交换与处理,因此ATM网络的传输时延较小.ATM采用面向连结的虚电路方式,虚电路由虚通道(VP)和虚通路(VC)两级组成.信元结构如图1所示.在信元头部中VPI与VCI分别是VP与VC的标识,每次通信前,由交换机进行路由选择,确定所选用的VPI及VCI值,建立端到端的虚电路连结.信元头部还包括一般流量控制(GFC)、净荷类型(PT)及信元丢失优

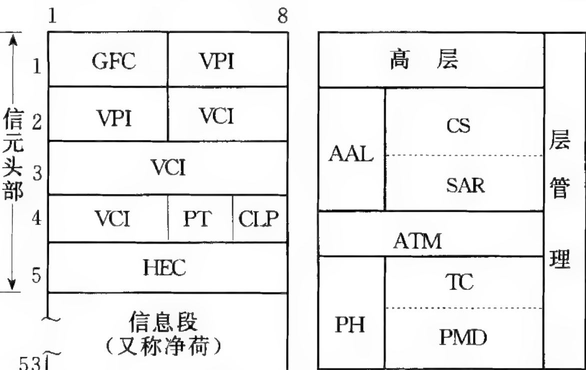


图1 ATM信元结构

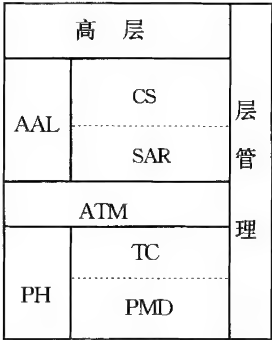


图2 ATM网协议分层结构

先级(CLP)和信头差错控制(HEC).ATM网络是由ATM终端(TE)及ATM交换节点组成.ATM网络协议分层模型如图2所示,分为物理层(PH)、ATM层、ATM适配层(AAL)以及高层.物理层又分为传输会聚子层(TC)及物理媒体子层(PMD).物理层提供物理媒体接口、位同步定时、信元定界与填充、传输帧的产生与恢复、信元速率适配等.ATM层完成信元拆/装、一般流量控制、VPI/VCI转换与解释、信元复用/解复用等功能.AAL层负责不同业务信息的适配,包括信息的分段与重装,组成AAL层协议数据单元(PDU),提供与高层原语交互接口等.AAL层分为会聚子层(CS)与拆装子层(SAR).为了适应不同类型业务传输的需求,AAL对业务进行了分类,ITU-T I.362定义了AAL的基本原理及功能类别,并根据业务的属性,将全部业务划分为四类:

1.面向连结的恒定比特率要求定时的业务,例如64kb/s的话音业务或固定比特率图象业务.

2.面向连结的可变比特率要求定时的业务,例如经过压缩编码(如MPEG)的可变比特率图象和语音.

3.面向连结的可变比特率业务不要求定时的业务,例如在分组交换网或帧中继网上的数据传输.

4.无连结型可变比特率不要求定时的业务,例如局域网业务、交换型多兆比数据业务或IP业务等.

迄今ITU-T已提出四种AAL协议,即AAL1,AAL2,AAL3/4和AAL5,分别适用于上述的1,2,3和4类业务.各类AAL协议的传输帧格式不同,差错控制方式及纠错编码也相异,它们所提供的服务质量(QoS)也呈现不同的特性.近年来,ATM规范不断完善,ITU-T,ANSI以及ATM论坛对ATM标准作了大量工作,推动了ATM技术的实用化进程.ATM交换机、接入设备、ATM专用处理芯片及网络适配卡等产品大量投入市场,并在网络中得到运用.但ATM技术复杂,产品价格昂贵,目前千兆位以太网,TCP/IP,SDH和DWDM技术在局域网与广域网形成对ATM的竞争格局,使ATM应用的发展受到限制.

电文交换(message switching) 存储转发交换的一种形式.一般是利用计算机来实现的.发信端用户首先把要发送的信息编成电文,连同收信端地址等辅助数据一起发给本地交换中心,本地交换中心将这些信息完整地存储起来并作适当处理,当需要的输出口有空时,就将电文转发到下一个交换中心,最后由收信端的交换中心将电文传送给用户.

分组交换(packet switching) 数据交换的一种方式.它把数据信息分割成若干个分组,每个分组除数据信息外还包含地址信息和分组编号,分组的长度可以固定,也可以根据需要变化.每个分组在网络中都以存储转发方式独立地进行交换.同一个信息产生的不同分组在网络中的传输路径可以不同.网络节点进行差错检验、重发和流量控制等操作.收

信端收到分组之后，按照编号将它们组合，还原成原始数据信息.

空分交换(space division switching) 一种交换方式. 即每一通道(传输通路)都由一单独的导体构成的交换方式. 空分交换机一般指的是它的接续部分采用空间分割方式，每个用户占有一定的空间位置，各用自己独用的导线. 用户之间的接续，均在相应空间位置的某些交换接点上进行. 一条被连通的电路在一定通信时间只能为一对用户服务.

时分交换(time division switching) 亦称时分复用交换. 一种交换方式. 电路接续设备利用时间分割方式进行多路复用时采用的交换方式. 按照这种方式，许多用户合用一公共传输通道，他们按一定的时间顺序轮流占用，每一用户在公共通道的循环周期中只占用其中一个时隙.

中继方式(trunking) 通信理论术语. 指电话交换网或电话局所属电话交换机及其线路系统的总体结构. 对于一个电话交换网来说，它主要包括：

1. 整个网的编号制度.
2. 局间连结方式和中继线数量.
3. 各局所用交换机制式和主要机件数量.
4. 长途电话、市郊、农村电话等的配合.

对于一个电话局来说，它包括：

1. 所用交换机制式.
2. 各种用户的分解计划和各用户之间的连结方式.
3. 主要机件数量和接续方式.
4. 与市话网和长途电话、市郊、农村电话等的配合.

话务量(traffic) 亦称电信负载量. 电信业务流量. 它表示电信设备承受的负载，也表示用户对通信需求的程度. 话务量的大小取决于在一定时间内所发生的呼叫次数和每次呼叫占用的时间. 通常，话务量是指在最繁忙时间内一小时的平均话务量. 单位时间内的话务量又称话务量强度，或话务负荷强度. 其值为一定数量的机线设备在单位时间内所承担的呼叫总次数与每次呼叫的平均占用时间之乘积. 另外，每个电话用户的平均话务量称为每线话务量. 一组负荷源的话务量指该负荷源在一定时间内(一天或一小时)发生的平均占用次数与每次占用时间之积，即

$$A = NCt,$$

式中 A 为话务量；N 为负荷源数(话务量来源数)；C 为一定时间内每个负荷源的平均占用次数；t 为平均每次占用时间. 话务量有几种不同的单位表示："小时呼"表示一小时内一组线路总共被占用的时间，"爱尔兰"表示一条电路的话务量. "小时呼"和"爱尔兰"在话务计算中是等价的.

集中控制方式(centralized control) 程控交换机所采用的一种控制方式. 它指整个交换机所有的控制功能，包括呼叫处理和维护管理功能，都集中在一个处理机上(这个处理机常称为中央处理机). 为保证交换机的可靠性，不能因控制设备故障而造成全局通信中断，因此，中央处理机必须采用冗余技术，如双机或多机同步方式、热备用方式、负荷分担方式等.

分布式控制方式(distributed control) 程控交换机所采用的一种控制方式. 它指系统采用多处理机，以一定的分工方式来承担整个交换机的控制功能(包括呼叫处理功能和维护管理功能). 处理机之间的分工可分为功能分担方式、负荷分担方式、容量分担方式以及三种分担的混合方式. 新一代的程控数字交换机大多数采用分布式控制方式.

呼损(call loss) 通信理论术语. 是衡量电话通信服务质量的一个指标. 因交换网的设备不足而使一些用户的呼叫不能接通的情况称为呼损. 这些不能接通的呼叫数与发生的总呼叫数之比称为呼损率.

呼损率(call loss rate) 见"呼损".

信令系统(signalling system) 通信网中为用户建立接续和进行网络管理采用的一种信息交换系统. 通信网为了控制通信接续的建立和监视接续的状态，在用户与交换机之间、交换机与交换机之间必须传送占用、应答、拆线、被叫电话号码等信息，这些信息称为信令. 传送信令的系统称为信令系统. 信令系统犹如通信网的神经系统，在用户和网络节点之间传递和交换控制信息. 按照传送信令的信道与话路的关系，信令系统分为随路信令和公共信道信令两种方式.

信令(signalling) 见"信令系统".

随路信令(channel associated signalling) 一种信令. 利用话音通络传送的信令称为随路信令. 由机电式交换机组成的电话网，一般采用随路信令方式. 随路信令分为线路信令和记发器信令两类，其国际标准是 ITU-T4 号和 5 号信令，中国标准是中国 1 号信令. 随路信令速度慢，信号种类少，只能满足基本电话业务的需要.

公共信道信令(common channel signalling) 一种信令. 在专用的信道上传送，用于控制一群话路的信令称为公共信道信令. 它是在程控交换技术发展的基础上产生的，将话路和信令信道分开，即话路信道只传送语音，而将很多话路的控制信令集中放到一条专用的高速数据链路上进行传送. 公共信道信令的国际标准是 ITU-T 6 号和 7 号信令. 这种信令传送速度快，信号种类多，适用于数字电信网.

7 号信令(NO. 7 signalling) 一种公共信道传

送信令. 是 ITU-T 建议的标准公共信道信令方式. 中国也制订了中国 7 号信令方式技术规范,7 号信令主要用于电话的局间信令、电路交换数据网的局间信令、ISDN 的局间信令、智能网中交换局与业务控制点之间的信令、用户小交换器(PABX)和交换局之间的信令,以及各种操作维护管理(OAM)中心的信息传递业务. 7 号信令系统的功能结构分为公用的消息传递部分(MTP)和适合于不同应用的多个用户部分(UP). MTP 提供用户部分之间可靠的消息传递功能,而每个用户部分是 MTP 的应用功能实体,各有其特定的功能. 7 号信令网由信令点、信令转换点和信令链路构成,独立于信息网存在,其工作方式分为直联、准直联和非直联三种.

编号方案(numbering plan)　对电话用户的号码安排. 要实现电话接续的自动化,必须使每一部电话机有一个电话号码,在电话网中对整个电话用户的号码安排方案称为编号方案. 编号方案的确立,应考虑以下因素:

1. 要远近期结合,即考虑近期的具体情况,又要考虑远期发展的需要,一般要考虑几十年的发展.

2. 尽可能缩短号长,具有规律性,以使用户使用方便.

3. 编号方案与交换系统的路由选择、计费等功能有密切关系,应尽可能使交换设备简单,以节省投资.

4. 在任何地点呼叫同一用户均拨相同的市话号码,不能因主叫用户所在地不同而有所变动.

5. 国内号码长度应符合 ITU-T 所制定的国际编号方案中最长位数的规定. 通常对于长途电话网,国际上一般采用固定编号方案. 如在原市话号码的前面加一个长途局号和一个用于与市话号码相区别的字冠.

光纤通信

光纤通信(optical fiber communication)　一种以光波载送信息,以光纤作为传输媒体的通信方式. 光纤(或称光导纤维)是一种由石英材料制作的介质传输线. 强度调制和直接检波是光纤通信系统目前

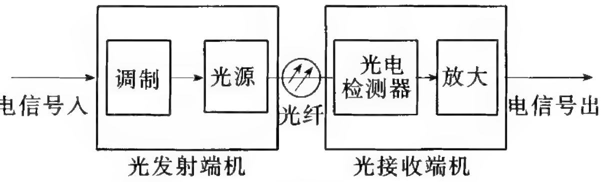


图1　光纤通信系统基本结构图

通常采用的技术,其基本结构如图 1 所示. 它是由光发射端机、光纤(光缆)和光接收端机组成. 图中的电信号可以是模拟信号,也可以是数字信号.

光源(发光二极管或半导体激光器)通过调制器将输入的电信号变成相应的光信号送入光纤. 光纤作为传输媒体以极小的衰减传送已调光信号. 光电检测器对输入的光信号进行直接检测,将其转变为相应的电信号再用均衡、放大等手段以弥补线路传输过程中的波形畸变和衰耗,最后输出和原始信号一致的电信号,从而完成通信的全过程. 在长距离通信中需在光纤中接入光放大器,将衰减后的光信号不断加以放大.

分析光纤的导光原理采用了两种方法:一为射线理论法;二为波动理论法.

由此原理导出光纤的两个重要参数:

1. 数值孔径 NA. 光纤是由一个拆射率(n_1)较大的纤芯和折射率(n_2)较小的包层组成,阶跃折射率光纤的数值孔径 NA 为

$$\mathrm{NA} = n_1\sqrt{2\Delta};$$

$$\Delta = \frac{n_1 - n_2}{n_1}.$$

它描述光纤接收光源射线的能力. 为了让光源功率更有效地耦合到光纤中,希望 NA 尽可能大,但这样会增加光纤的模式色散. ITU-T 建议 NA 的取值范围在 0.18～0.24.

2. 归一化截止波长 λ_c. 光纤中波导模式 $\mathrm{LP}mn$ 截止时的 λ_c 值如表 1 所示.

表 1　几种 $\mathrm{LP}mn$ 模的 λ_c 值

n \ m	0	1	2
1	0	2.40483	3.83171
2	3.83171	5.52008	7.01559
3	7.01559	8.65373	10.17347

进入光纤中的光信号具有一定归一化频率 ν,合理选择光纤纤芯半径 a 和 n_1,n_2,可使

$$0 < \nu = \sqrt{2\Delta}n_1 \cdot \frac{2\pi}{\lambda_0}a < 2.40483.$$

从而使光纤中只传输 LP_{01}模式,这就是阶跃光纤单模传输的条件,信号波长 λ_0 越长,a 的尺寸越可以增大,有利于单模光纤的制作和使用.

单模光纤的主要传输性能是损耗和色散,它直接影响光中继传输距离和传输带宽. 其含意是:

1. 损耗. 损耗常用衰减系数 A 表示. 如在长 L 公里的光纤上有输入功率 P_i、输出功率 P_o,如损耗均匀,则

$$A = -(1/L)\lg\frac{P_o}{P_i}.$$

引起光纤损耗的机理非常复杂,造成光纤损耗的原

因简单归纳于表2中.

表2 光纤损耗原因

名 称	原 因
吸收损耗	本征吸收
	杂质吸收
	原子缺陷吸收
散射损耗	瑞利散射
	结构不完善引起散射
弯曲损耗	光纤弯曲

损耗的主要特性与光纤工作波长 λ 有关,图2是光纤损耗谱曲线. 图中表明了波长为 1.31μm 和 1.55μm 附近是两个低损耗窗口. 目前新的技术已消除了 1.4μm 波长左右的衰减峰值. 因此 1.3～1.6μm波段均可用来通信.

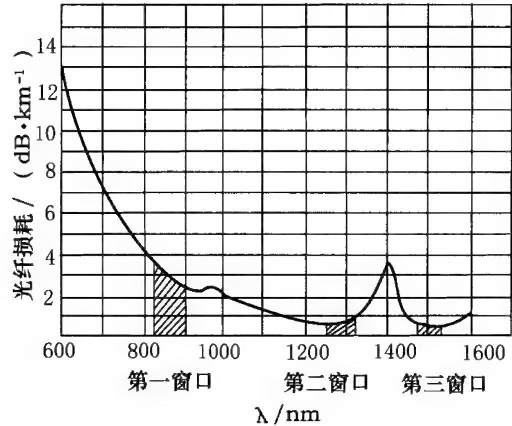


图2 三个常用的光纤低损耗窗口

2. 色散. 光纤色散是由于光纤传输信号的不同频率成分、不同模式成分的群速不同而引起时延差,使传输波形畸变的一种物理现象,时延差的单位是 ps/(nm·km),时延差又称为色散系数. 色散限制了中继距离和通信容量. 单模光纤包含材料色散与波导色散,典型的单模光纤色散与波长的关系曲线如图3所示. 从图中可见,在 λ=1.27μm 附近材料色散为零;在 1.3～1.4μm 之间材料色散与波导色散相互抵消为零. 综观损耗与色散曲线,使零色散点移至 1.55μm 附近则可获得最低损耗和最小色散,即 ITU-T 建议的 G.653 色散位移单模光纤. 而原来在 1.3μm 波段色散为0的光纤称为常规光纤(G.652). 由于 G.653 光纤不适宜采用当前国际上普遍采用的波分复用技术(参见"波分复用技术"),在1.5μm波段色散范围内,但色散不为0的非零色散位移光纤(G.655)便得到了很大发展. 目前可以说G.652光纤是世界上铺设最多的光纤. G.655 光纤是最有发展前途的光纤. 在传输速率等于或超过 40GB/s 时,偏振模色散(参见"偏振模色散")也是必

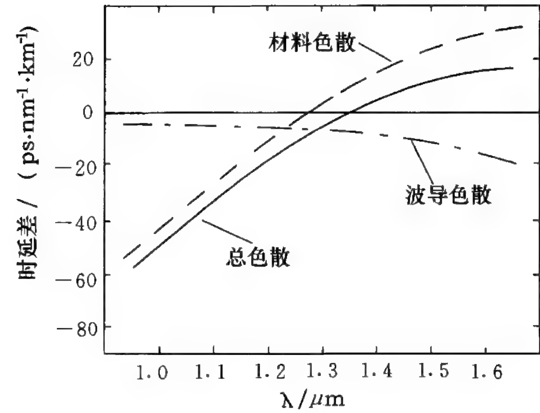


图3 单模光纤色散与波长关系曲线

须克服的色散之一.

同步数字体系(synchronous digital hierarchy)简称 SDH. 一种传送信号的体制. 即继点对点通信的准同步数字体系 PDH(pseudochronous digital hierarchy)之后,向以光纤为主要传输媒体的同步数字网(SDH 网)发展的一种新型的数字通信体制. SDH 的标准传输速率如下表所示.

SDH 的标准传输速率

SDH		SONET	
等 级	速 率 /(MB·s⁻¹)	等 级	速 率 /(MB·s⁻¹)
		STS-1	51.840
STM-1	155.520	STS-3	155.520
		STS-9	466.560
STM-4	622.080	STS-12	622.080
		STS-18	933.120
		STS-24	1244.140
		STS-36	1866.240
STM-16	2488.320	STS-48	2488.320
		STS-96*	4976.640
STM-64*	9953.280	STS-192*	9953.280

注:1. * 表示尚待研究.

2. 美国贝尔通信研究所提出的同步光网络(SONET)是北美标准,SDH 是在此基础上发展起来的.

SDH 网的一个关键功能是要求包括 PDH 支路信号,能够进行数字复用、交叉连结和交换,因而帧结构应便于对支路信号接入和取出. ITU-T 根据这些要求建议了如图1所示的帧结构. SDH/SONET 中采用了净负荷指针技术,可以避免采用 125μs 缓存器和在复用设备接口的滑动,指针指示了净负荷在 STM-N 帧内第1个字节的位置,净负荷在 STM-N 帧内是浮动的. 对于净负荷的较小频率变化只需增减指针值即可.

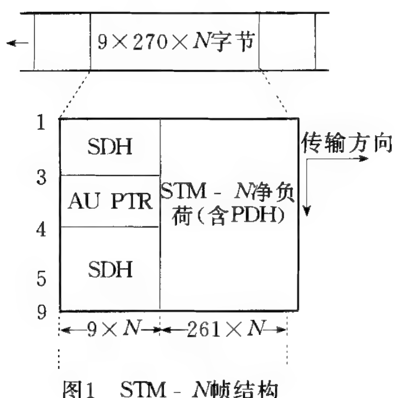


图1　STM - N帧结构

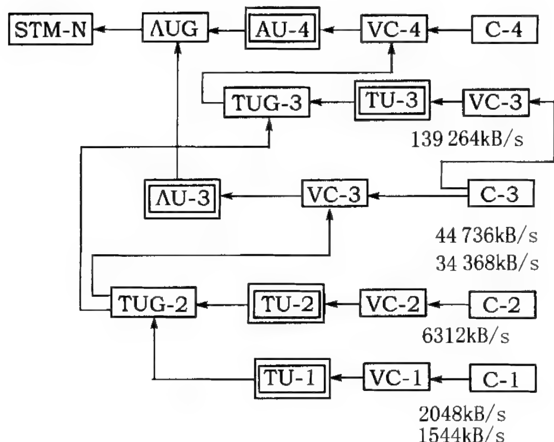


图 2　SDH 的复用过程

SDH 的具体复用过程如图 2 所示.

SDH 网络单元(ADM 或 SDXC)作为网络节点连成一个环形网并构成自愈网,是 SDH 的重要特色,极大地增强了网络的生存性.

光纤(optical fiber)　亦称光导纤维.一种可以传导光的介质.20 世纪 70 年代以来,光纤的损耗大幅度下降,目前损耗已低于 0.2dB/光路,因而可以用它长距离传输光信号.用于制作光纤的材料主要是 SiO_2.塑料光纤由于损耗大,仅用于短距离通信或传感器.还有一些特殊材料的光纤,如氟基质光纤等.从传输特性看,光纤可以分成单模光纤、多模光纤、色散位移光纤、非零色散位移光纤、色散补偿光纤和保偏光纤等.

光导纤维(optical fiber)　即"光纤".

光纤色散(optical fiber dispersion)　光在光纤中传播的一种特性.不同波长或不同模式的光在光纤中传播速率不同,称此为光纤色散.不同波长或不同模式的光在光纤中传播的群速度不同称为群(速度)色散.由于材料不同而引起的色散称为材料色散;由于波导结构的存在而引起的色散称为波导色散.上述两者统称色度色散.由于不同传输模式引起的色散称为模式色散.特别地,由于偏振模不同引起的色散称为偏振模色散(PMD).光纤的色度色散常用一阶群色散$\ddot{\beta}$或色散系数 D 表示:

$$\ddot{\beta}=\frac{\mathrm{d}}{\mathrm{d}\omega}\left[\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\right]=-\frac{1}{v_g^2}\frac{\mathrm{d}v_g}{\mathrm{d}\omega},$$

$$D=\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\right]=-\frac{2\pi c}{\lambda^2}\ddot{\beta},$$

其中 β 为光纤的传输常数,ω 为光波圆频率,λ 为光波波长,c 为光速,v_g 为光纤中光传输的群速度.按照光学的习惯,当$\ddot{\beta}>0$ 时称为正常色散,当$\ddot{\beta}<0$ 时称为反常色散.

偏振模色散(polarization mode dispersion)　亦称极化模色散.简称 PMD.光在光纤传播中的一种特性.指光纤芯层中两个空间正交的简并传播模式之间的传播群速度差.在光纤的芯层中,由于波导结构和边界条件的对称性,其内光波以两个空间正交的简并模形式传播.在理想情况下,对于弱导光纤两个正交简并模具有相同的群速度(或称具有相同的群时延),不存在偏振模色散.在实际的光纤中,由于各种因素会导致沿光纤轴向出现波导结构和边界条件的对称性破损,其中主要包括几何圆度的扭曲、光纤的扭转、折射率分布的角向不均匀性等,从而引起光纤内的双折射现象,使得两个原本空间正交的简并模的群速度出现差异,使光纤中传播的光波发生极化模色散.

一般情况下,极化模色散与光波的频率(波长)有关.在实际的系统中,其变化具有时间上的随机性,故多采用统计均方根值来定量描述.按照它与频率(波长)的变化关系可以分为一阶偏振模色散和高阶偏振模色散.

极化模色散(polarization mode dispersion)　即"偏振模色散".

受激拉曼散射(stimulated Raman scattering)　简称 SRS.光纤通信术语.受激拉曼散射和受激布里渊散射(SBS)都是光纤中的非线性参量.它起因于光场与介质相互作用的过程中的受激非弹性散射,即在这一过程中光场将部分能量传递给介质.其量子力学图象为入射泵浦光的一个光子的湮灭,同时产生一个频率下移的斯托克斯光子和一个(与介质振动激发态有关的)声子.为了保证整个过程中的能量和动量守恒,所产生的声子应具有恰当的能量和动量.两者的主要不同在于:SRS 产生的声子频率比较高,称为光学声子,而 SBS 产生的声子频率比较低,称为声学声子.在光纤传输中的一个根本差别是 SBS 只发生在后向,而 SRS 则主要发生在前向.

受激布里渊散射(stimulated Brillouin scattering)　见"受激拉曼散射".

光学声子(optical phonon)　见"受激拉曼

散射”.

声学声子(acoustic phonon)　见“受激拉曼散射”.

自相位调制(self-phase modulation)　简称SPM.光纤通信术语.指光纤中信号传输的一种特性.由于光纤的折射率具有非线性特性,当光纤中电场强度变化时光纤的折射率随着变化,光纤中传输的信号相位也变化.信号自身场强的变化引起自身相位的变化,这就是自相位调制.自相位调制在光纤通信中最重要的应用是它和光纤色散的相互作用可以在光纤中实现光孤子传输.

交叉相位调制(cross-phase modulation)　简称CPM.光纤通信术语.指光纤中信号传输的一种特性.交叉相位调制与自相位调制产生的机理是类似的.光纤中一个信号的场强变化引起另一信号的相位变化称为交叉相位调制.交叉相位调制可能引起光纤中信道之间的串扰,这是应当避免的.

四波混频(four wave mixing)　简称FWM.信号在光纤传输中产生的一种现象.光纤中的自相位调制、交叉相位调制和四波混频都是由光纤的三阶非线性极化率引起的.在四波混频过程中,一个或几个光子被湮灭,同时产生几个不同频率的新光子,在此过程中,净能量和动量是守恒的.例如,圆频率为ω_1和ω_2的两个光子湮灭,同时产生圆频率为ω_3和ω_4的新光子,并有

$$\omega_3 + \omega_4 = \omega_1 + \omega_2$$

和

$$\Delta k = k_3 + k_4 - k_1 - k_2 = 0,$$

其中k_i为对应圆频率为ω_i的波数.

四波混频过程要求满足相位匹配条件,上述第二个式子即代表相位匹配条件.

光放大器(optical amplifier)　一种光学器件.是放大光信号的设备或器件的总称.它又分为光纤放大器、拉曼放大器和半导体光放大器等.它们具有不同的工作原理和用途.

光纤放大器(optical fiber amplifier)　光放大器的一种.20世纪60年代开始提出,20世纪80年代末开始应用.它的使用给光纤通信带来了质的飞跃.它是使用稀土离子作为增益介质,如将铒离子(Er^{+3})掺入纤芯制成掺铒光纤而提供光增益的光纤放大器,称为掺铒光纤放大器(EDFA).在泵浦光的作用下,掺杂离子能级从基态升至激发态.由于激发态的不稳定,在离子从激发态返回基态过程中可直接对光信号进行放大.

光纤放大器的主要参数有放大增益、饱和输出功率、放大带宽、噪声系数和工作波段等.一般光纤放大器在工作通带内增益和输出功率不同,对用于波分复用光通信系统中的光纤放大器,还要有增益锁定和功率均衡功能,最近不少放大器采用了瞬态增益控制技术.由于光纤放大器存在放大器自发辐射(ASE)噪声,光纤放大器在级联时光信号信噪比将会降低.

光纤放大器从功能上可分为前置放大器、功率放大器和在线放大器三种.前置放大器要求有低噪声指数(NF),增益要求不很高;功率放大器要求信号饱和输出功率大或增益要大,对NF要求不高;在线放大器要求低NF,但增益要大.

光纤放大器已取代传统电中继放大器,并增加了光中继之间的信号传输距离.在N波长信道波分复用光纤通信系统中,一个光纤放大器可取代N个或$2N$个传统电中继器.

光纤拉曼放大器(optical fiber Raman amplifier)　光放大器的一种.由于光纤的非线性,当光通过光纤传输时将产生受激拉曼散射(参见“受激拉曼散射”).光纤拉曼放大器就是利用受激拉曼效应将强泵浦光能量转移到信号光束上,从而达到放大光信号的目的.

光纤拉曼放大器的主要优点有:只要能得到所需的泵浦波长,可以为任何波长提供增益及低噪声.因此现有将光纤拉曼放大器置于光纤放大器之前的用法,可使放大器之间的传输距离更长.它的主要缺点是:需高泵浦和泵浦偏振等问题.

半导体光放大器(semiconductor optical amplifier)　简称SOA.光放大器的一种.半导体光放大器同光纤放大器一样由激活介质组成,只不过半导体光放大器是一个具有或不具有端面反射的半导体激光器.电流使介质的载流子反转到激发态,从而使外输入光产生受激辐射而获得相干增益.半导体光放大器增益谱带宽大,约70nm左右,对偏振敏感,由交叉饱和引起的信道间串扰问题严重,而且尾纤耦合损耗大.通常利用半导体光放大器的放大增益和非线性特性来制作全光波长变换、光子交换门等.

放大器自发辐射噪声(amplifier spontaneous emission noise)　简称ASE噪声.光纤通信术语.指在光纤放大器和半导体光放大器中,随着激活粒子从激发态返回基态并放大光信号的同时,也会产生受激粒子的随机非相干自发辐射.这种自发辐射可在任何方向,并可引起进一步受激辐射,且可被放大.它的频带很宽,可占据整个增益带宽.称此为放大器自发辐射噪声.在上述光放大器中,在一定条件下被激发的粒子数是确定的,被用于产生ASE的粒子数愈多,可用于提供信号增益的粒子数目也就愈少.

光纤通信系统(optical fiber communication system)　一种新型通信系统.即以光纤作为信息传

输媒质的系统.光纤通信系统通常包括发射机、光放大器和接收机等.早期光纤通信系统每芯光纤仅传输一个波长,所有信息均在电域处理,光仅仅作为信息传输中的载体.在第二代光通信发展过程中,部分信息已在光域处理,如光合波、分波、光交换等.将来在信息传输过程中,除了发送端和接收端可在电域处理外,中间过程全在光域中进行处理.

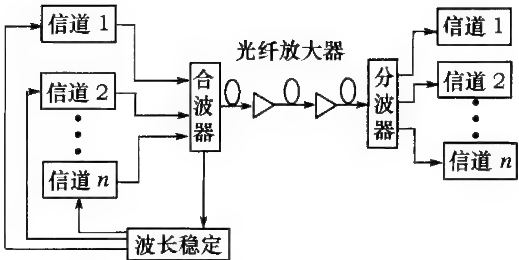
目前已经可能利用光纤建设电信网、电视分配网和计算机网.光纤通信网将成为信息基础设施的骨干网.

时分复用技术(time division multiplexing technique) 简称TDM技术.光纤传输中采用的一种复用技术.将不同数字信号序列在时域中按一定规则占有不同的时隙所组合成一个新的数字信号序列的技术.通过时分复用技术,若干不同的低传输速率的传输信号组合成了高传输速率的传输信号.在光通信领域中时分复用技术可分为电时分复用和光时分复用.电时分复用就是若干不同的低传输速率的数字电信号在电域中复用成新的高传输速率的数字电信号后去调制光源,产生光信号序列;而光时分复用是具有相同波长的若干不同的低传输速率的数字光信号序列在光域中复用成新的高传输速率的光信号序列.由于目前器件材料的限制,电域最高调制传输速率已接近器件的可使用极限.要继续提高单路光信号的传输速率,光时分复用技术是惟一的选择.

电时分复用(electric time division multiplexing) 见"时分复用技术".

光时分复用(optical time division multiplexing) 见"时分复用技术".

波分复用技术(wavelength division multiplexing technique) 简称WDM技术.一种光纤通信技术.波分复用技术是指在一芯光纤中传输多路不同光波长的技术.采用波分复用技术的通信系统或网络通常称之为光波分复用通信系统或波分复用通信网.相邻波长间隔小(通常为几十GHz到数百GHz)的波分复用通信系统称为密集波分复用(DWDM)通信系统.同一波分复用通信系统中每个波长上的调制信号可以是模拟信号或各种制式的数字信号.波分复用通信系统至少包括以下几个部分:不同波长光发送源组、合波器、光纤放大器、光纤、分波(解复用)器和接收机组.为保证系统正常工作,通常还需要对光发送源中激光器波长进行稳定与控制.下图是波分复用通信系统原理图.



波分复用技术虽然很早就提出来了,但直到20世纪80年代末光纤放大器商品化后,该技术才得到应用与推广.波分复用技术是扩大光纤传输容量的最佳办法.2001年通过波分复用技术单芯光纤传输容量已达10.92TB/s(273波长×40GB/s).

光波分复用通信系统(optical wavelength division multiplexing communication system) 见"波分复用技术".

密集波分复用通信系统(dense wavelength division multiplexing communication system) 见"波分复用技术".

光信噪比(optical signal-to-noise ratio) 光纤通信术语.即在一定带宽范围内光信号功率与光噪声功率的比值.通常是指在光有效带宽为0.1nm内光信号功率和噪声功率的比值.虽然光信噪比并不能表征光信号本身的恶化程度,但它包含误码信息,直接影响接收机端的噪声功率水平.因此,在光通信系统中,光信噪比是一个十分重要的参数,对估算和测量系统的性能有重大意义.由于难以分别测量在同一带宽范围内光信号与光噪声的功率,因此,实际中往往通过光谱形状对光信号和光噪声的功率进行估算.光通信系统中的噪声主要来源于光放大器的ASE噪声(参见"放大器自发辐射噪声")积累,信号每通过一级放大器均伴随着光信噪比的恶化.

量子噪声(quantum noise) 光纤通信术语.指光的粒子性产生的噪声.在光通信中,光到达接收机时将以光量子的形式随机到达,单位时间到达的量子数满足泊松分布.由于到达的量子数的随机起伏,所产生的光电流也随机起伏,这一起伏就是量子噪声.由于它是光的粒子性产生的,因此又称为光散粒噪声.

光散粒噪声(optical shot noise) 见"量子噪声".

功率均衡(power equalization) 网络理论术语.指波分复用光网络中各个波长信道间功率水平由于各种原因可能会存在比较大的差异,这种差异既可能来自于静态的波分复用传输,也可能是来自于动态的光变换.在波分复用传输系统中,功率不均衡主要来源于光放大器增益非平坦特性及传输光纤的非线性效应.在光网络中引入光交叉连结(OXC)节点(参见"光交叉连结")和光上下分插复用(OADM)节点(将光分插复用器)后,不同波长信道在节点内部经过的路径不同,解复用器(参见"解复用器")、光开关(参见"光开关")对它们引入的损耗

也不相同,此外网络的动态重构会进一步引起各个波长信道功率水平的起伏.

功率不均衡将会对系统的性能造成比较大的损害,限制网络的规模和容量,因此需要对各个波长信道的功率水平进行调节,使之基本保持一致,这就是功率均衡需要完成的任务.对于大规模的光网络,节点内部一般都有专门的功率均衡模块来调整各个波长信道的功率水平.

复用器(multiplexer) 光纤通信系统中的一种器件.实现将多芯光纤中不同波长的光信号合入一芯光纤中的器件称为复用器;若将一芯光纤中的多路光信号分离至多芯光纤中的器件则称为解复用器.复用器和解复用器可由光波导、光纤、体光栅或薄膜等构成.

解复用器(demultiplexer) 见“复用器”.

波长转发器(transponder) 光纤波分复用系统中的一种器件.将某一光波长上负载的信息转移到另一所需光波长的器件.有光/电/光和全光两种形式.光/电/光形式实际上是一对光电收发端机,采用了电 3R(再生、整形和时钟提取)技术,它仅是目前在全光器件未成熟之前的一种过渡性器件.

采用了波长转发器的波分复用通信系统成为开放式系统,可接入非波分复用通信系统的信息;光波分复用网络节点中采用了波长转发器后,将可实现虚波长路由,降低网络阻塞率.

光波导(optical waveguide) 光纤通信术语.指光能流被限制在其内或其表面层附近传输的任何结构.

光纤布拉格光栅(optical fiber Bragg grating) 一种光纤系统传输器件.指利用某些光纤紫外光敏特性使纤芯折射率沿轴向形成小的周期性调制而形成的光纤器件.折射率变化在 10^{-5}~10^{-3}之间.制作方法有相干法、相信模板法、逐点写入法等.

光纤布拉格光栅分透射和反射两种类型.对于透射式,当两个同向传播的波的传输常数满足如下相信匹配条件时:

$$|\beta_0 - \beta_1| = \frac{2\pi}{\Lambda},$$

一个波的能量会耦合进另一个波中.其中,Λ 为光纤折射率调制周期,β_0,β_1 分别为两个波的传输常数.在反射式光栅中,入射波与反向散射波耦合,相位匹配方程为

$$|\beta_0 - (-\beta_0)| = 2\beta_0 = \frac{2\pi}{\Lambda},$$

令 $\beta_0 = 2\pi n_{\text{eff}}/\lambda_0$,$\lambda_0$ 为入射波的波长,n_{eff}为光纤的有效折射率,则波长满足下式的波被反射:

$$\lambda_B = 2n_{\text{eff}}\Lambda,$$

λ_B 为布拉格波长,当入射波长与布拉格波长失谐时,反射率将变低.

下列图中是几种光纤光栅类型:

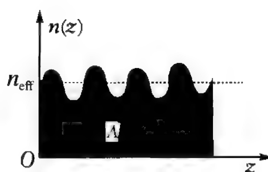


图 1 纤芯折射率分布示意图

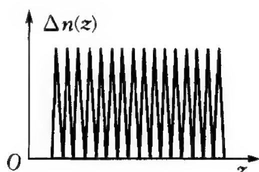


图 2 均匀光栅

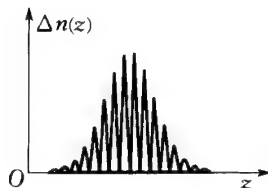


图 3 高斯切趾光栅

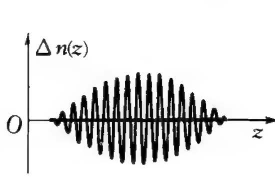


图 4 余弦切趾光栅

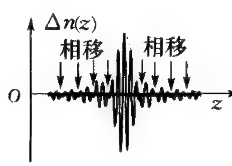


图 5 sine 切趾光栅

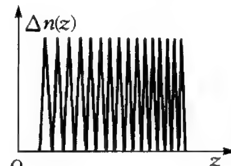


图 6 啁啾光栅

啁啾光栅(chirped grating) 一种常用光栅.啁啾光栅的折射率变化为

$$n(z) = n_{\text{co}} + \overline{\Delta n(z)}\left\{1 + \cos\left[\frac{2\pi}{\Lambda}z + \Phi(z)\right]\right\}.$$

啁啾光栅常用作波分复用通信系统中的色散补偿.

非线性光环镜(nonlinear optical loop mirror) 简称 NOLM.一种光学装置.指在光纤环形光路中含有非线性光学介质的 Sagnac 干涉装置.其中的非线性光学介质可以是一段光纤,也可以是一只半导体光放大器(SOA)或其他非线性光波导.最简单的光环镜结构只需将一只 2×2 的 3dB 单模光纤光耦合器的两个输出端口焊接起来即可.它有多种应用,其中包括全光开关、全光解复用器、全光信号再生和时钟提取以及逻辑运算等.

相干光纤通信系统(optical fiber coherent communication system) 光纤通信系统的一种.指利用激光器发出的光的相干性,在接收端实现外差或零差检测,称为相干光纤通信.利用该方法实现信息传送的通信系统称为相干光纤通信系统.该系统可以通过振幅、频率、相位或偏振态来传递信息.

相干光纤通信系统的优点是接收灵敏度和频率信道分辨率高.灵敏度一般比直接调制高 20dB 左右.灵敏度同本振功率有关.当本振功率足够高时,相干系统只受接收信号量子噪声的限制.另外相干

光纤通信系统信道间隔小，可实现的信道数多，选择性好. 但它对光源的要求高(要求光源的线宽很小)，接收信号和本振信号的偏振应相互匹配.

频分复用技术(frequency division multiplexing technique) 简称 FDM 技术. 一种光纤通信技术. 是系统复用技术的一种. 当多路复用通信系统中的相邻波长间隔更小时(通常小于 1GHz 以下)，该系统称为频分复用通信系统. 频分复用技术通常又分为相干和非相干两种类型. 非相干频分复用系统从系统结构上类似于波分复用系统，但它对频率(波长)稳定性要求更高，且调制方式不采用幅度调制，常采用频率调制. 相干频分复用系统通常还要采用频率跟踪等电路，以确保有稳定的输出中频. 该系统采用的器件要求同波分复用系统也不一致. 因此，频分复用技术同波分复用技术可以说是完全不同的复用技术.

光孤子通信(optical soliton communication) 光纤中的一种光脉冲无畸变传输的正确方式. 光纤中的光孤子是利用光纤中的色散和非线性效应的相互补偿作用实现光脉冲无畸变的传输. 无论在光纤的正常或反常色散区，光脉冲在传输过程中都有脉冲展宽现象. 光纤的自相位调制(非线性)效应的单独作用也会使脉冲展宽. 但是，在反常色散区，色散和自相位调制的作用可以相互抵消，使光脉冲在传输过程中脉冲宽度保持不变. 如果光纤传输过程中有能量补偿光纤的损耗，则脉冲的幅度和宽度都可能不变化，从而形成光孤子传输. 在理想情况下，光脉冲可以在光纤中无限制地无畸变传输. 但是由于光纤放大器的引入，产生了附加的噪声. 引起了接收信号的时间抖动，限制了传输距离. 该限制称为戈登-豪斯极限. 一般认为这一极限值为 1 万公里. 为了克服这一极限，人们提出了频域滤波和时域滤波的方法，突破了这一极限，使光孤子在光纤中的传输距离达到数万公里.

戈登-豪斯极限(Gordon-Haus limit) 见"光孤子通信".

混沌通信(chaotic communication) 现代通信领域的一个新的分支和重要方向. 从 20 世纪 90 年代初，国际上相继提出了各种混沌通信制式及其理论与方法. 混沌通信主要划为四大类：混沌扩频、混沌键控、混沌参数调制和混沌掩盖. 前三类属于混沌数字通信，最后一类属于混沌模拟通信.

目前，对混沌通信的研究主要是在理想信道的范围内进行的，研究表明：混沌信号是类似高斯白噪声的宽带连续谱非周期信号. 利用理想信道(即有线信道)无失真地传送发端所有同步信息能较好地实现收发混沌的同步，然而，要实现混沌无线通信，就必须考虑实际信道、噪声干扰以及混沌信号高频载波调制和解调等的非线性失真对混沌同步的影响.

无线光通信(optical wireless communication) 利用自由空间作为传输介质的光通信. 它用光作为载波，实现信息的传输. 无线光通信主要应用于两种情况：空间(卫星间)光通信和地面无线光通信. 前者又分低轨道卫星之间和低轨道卫星与同步轨道卫星之间的光通信，传输距离最大可达 4 万公里. 其中卫星间瞄准、捕捉和跟踪技术是关键技术. 后者受大气环境的影响较大，一般只能传输几公里到十几公里.

异步转移模式(asynchronous transfer mode) 简称 ATM. 一种快速分组交换技术. 国际电信联盟 ITU-T 在 I.113 建议中给 ATM 以下的定义：它是一种传输模式，在这一模式中，信息被组织成固定长度的信元，而包含一段信息的信元并不需要周期性地出现，因此，这种转移模式是异步的. 每个信元含有 53 个字节，分为两个部分：前面 5 个字节为信头，主要完成各种功能；后面的 48 个字节为信息段，用来装载来自不同用户、不同业务的信息. 话音、数据和图象等所有的数字信息都要经过切割，封装成统一格式的信元在网中传递，并在接收端恢复成所需格式.

由于 ATM 技术采用硬件方式进行转发，简化了交换过程，去除了不必要的数据校验，所以 ATM 交换速率大大高于传统的数据网，如 x.25，DDN 和帧中继等. 另外，ATM 采用的是统计复用的技术，网络资源利用率很高. 但它的技术过于复杂，导致其应用困难.

光传送网(optical transport network) 简称 OTN. 网络的一种类型. 指在光域内实现业务信号的传送、复用、路由选择、监控，并且保证其性能指标和生存性的传送网络，它满足 ITU-T G.805 中规定的传送网的通用模型，遵循其分层结构、信息定义、客户关系、网络拓扑、功能界定等. 光传送网的层次结构自上而下依次为：电路(客户)层网络、光通道层网络、光复用段层网络、光传输段层网络和物理媒质层网络. 相对于 SDH 传送网(参见"同步数字体系")，光传送网有着诸如信号完全透明性、可扩充性、灵活可重构性、无电子瓶颈等诸多优点. 目前，光传送网发展的困难在于缺乏光域中完善的性能检测和故障管理机制. ITU-T G. 871 草案规定了关于光传送网的建议框架，G. 872 草案规定了光传送网的体系结构.

波分复用全光网(wavelength division multiplexing all-optical network) 采用波分复用技术的光纤系统网. 波分复用全光网由节点(光交叉连结和光分插复用器)和节点间的光纤链路的集合组成，通过网络管理对之进行维护和运行. 在网络中采用波分复用技术，并通过波长路由来完成信息的传送.

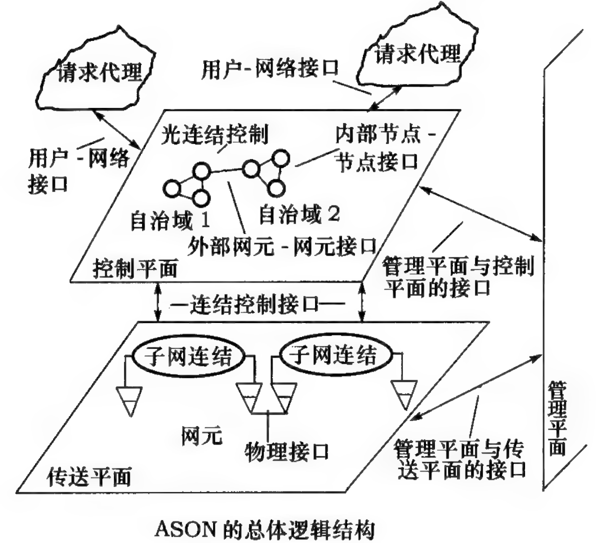
波分复用全光网具有透明性、容量大、可扩展性、兼容性、可重构性、高灵活性和强生存性等优点.可传送任何传输速率、任何制式的数字和模拟光信号.

由于目前光器件和技术的原因,如网络节点中采用光/电/光型波长转发器,破坏了网络的透明性.

自动交换光网络(automatic switched optical network) 简单 ASON. 网络转换的一种技术. 即国际电信联盟(ITU-T)研究和制定的智能光网络标准. 当接收到用户请求时,它能够在信令网的控制之下自动实现光网络内部连结,不需要网络管理员的人工干预.

自动交换光网络是在传统光传送网的基础上发展起来的,传统的光传送网主要依靠人工配置来实现连结的建立,难以适应快速开展新业务及市场竞争的需要. 而在自动交换光网络中则引入了动态交换的概念,支持电子交换设备(包括 IP 路由器、ATM(参见"异步转移模式")交换机及 SDH(参见"同步数字体系")设备等)动态向光网络申请带宽,这是传送网技术的一次重大突破. 自动交换光网络带来的好处主要有:

1. 实时的流量工程控制. 根据网络当前业务需求,动态分配和调度资源,使网络性能达到最优.



ASON 的总体逻辑结构

2. 灵活、高效的保护和恢复机制. 当网络发生故障时仍然能够维持一定水准的服务水平,特别是强大的分布式恢复能力可以有效地保证网络生存性.

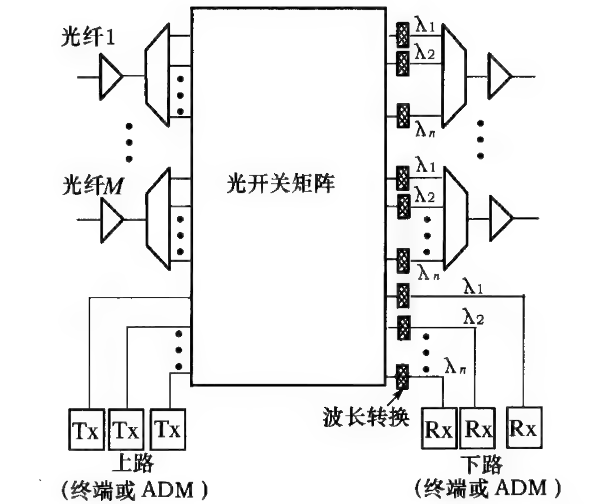
3. 能够快速、高质量的为用户提供各种带宽服务与应用,如光拨号、波长批发、波长出租等.

4. 良好的设备互操作性和网络可扩展性. 通过定义统一的、标准的网络接口,不同厂商的设备可以很容易的实现互联互通.

自动交换光网络的总体逻辑结构如下图所示. 整个网络涉及三个平面:传送平面、控制平面和管理平面.

光开关(optical switch) 实现光路交换的器件. 用来实现光路交换、光纤中光波长的重新组合分配、光网络的动态重构等. 从工作原理上划分,有机械光开关、微机电光开关、液晶光开关、热光开关、半导体光开关、声光开关等. 开关速度从毫秒量级到纳秒量级不等. 是光网络中的重要器件.

光交叉连结(optical cross connection) 简称 OXC. 光交叉连结是光通信网中重要的网络节点之一. 它将来自不同光纤中的复用波长在光域进行交换和重新组合. 结构如图所示.



光交叉连结有三种类型:

1. 没有光开关矩阵的固定波长路由节点,其不同光路波长的物理连结状态是固定的;

2. 具有光开关矩阵的但无波长转换器的节点;

3. 具有光开关矩阵的且有波长转换器的节点.

后两者具有动态波长交换能力,不同波长的物理连结状态可根据需要而改变.

光分插复用器(optical add drop multiplexer) 简称 OADM. 是光网络节点之一. 其功能是从传输设备中有选择地下达本地的光信号、及将本地光信号上路至传输设备中,同时不影响其他波长信道的传输. 光分插复用器可分为静态和动态两种. 在静态光分插复用器中,可上下路固定波长的光信号,而动态光分插复用器中可选择地上下路波分复用系统中一路至全部波长的光信号. 没有被选择上的上下路的光信号在该节点中经过处理(如经过光放大)后直接通过该节点,不需要经过光电变换处理.

光通路(optical path or light path) 光纤通信

术语.指在光网络的两个接入点之间的光层的点到点逻辑连结.信号在一条光通路内从入口到出口始终以光的形式传播,可以经过多个节点,也可以经过波长变换(参见"虚波长路由").在波长路由的光网络中,一条光通路是光层的最小业务单位,光层以提供光通路的形式向上层的电网络提供服务.在全光网络中(参见"波分复用全光网"),光通路具有对信号格式和传输速率等的透明性,可以在光域上进行交叉连结(参见"光交叉连结")和上下路(参见"光分插复用器")等处理.在没有波长转换的情况下,一条光通路就承载在一个光波长上,这就是波长连续性原则.在有波长转换的情况下,一条光通路可以跨越几段不同波长的链路.

网络生存性(network survivability) 网络理论术语.指网络在经受各种故障(例如线缆断裂、节点设备故障等)后能够维持可接受的业务质量的能力,它是网络完整性的一部分.网络生存性的研究内容主要包括建网初期对网络资源的优化配置和在网络运营过程中所采取的有效方式.目前,已有的提高网络生存性的方式有三种:自愈环、保护和恢复.在SDH(参见"同步数字体系")、ATM(参见"异步转移模式")和因特网协议(IP)等上层应用层中,对于网络生存性的研究已经较深入.随着波分复用光传送网的链路容量和传输速率的不断提高,光层上的网络生存性问题越来越重要.

自愈环(self-healing ring) 提高网络生存性的方式之一.它是一种主要用于环形网络中的提高网络生存性的方式.主要有以下几种基本结构:1+1线路倒换环、二纤单向线路倒换环、二纤双向线路倒换环和四纤双向线路倒换环.自愈环对于环形网络的链路故障具有很高的生存性,其业务的保护倒换时间很短,但是它所需要的空闲容量很大,一般情况下往往要求高达100%的备用容量,因此,自愈环结构存在着资源浪费的问题.另一方面,在无备用节点设备的网络中,自愈环结构对于节点故障是无能为力的.

波长路由(wavelength routing) 网络理论术语.指在波分复用技术(WDM)光网络中,节点之间的连结请求是用光波长来建立的,因此,信息的传送是通过透明的光路而无需经过光电转换.波长路由是指光信号在经过网络节点时,根据它的波长来选择路由.波长路由网络有两个显著的优点:

1. 由于波长决定了光信号传输的路径,因此,一个节点可以同时发出多路不同的波长信号,每路信号到达不同的目的地,目的地的数量与这个节点所能产生的波长数相同.

2. 由于每路信号被限制在特定的波长通道中,因此,只要这些波长通道不在同一条光纤中,在网络的其他路径中就可以同时使用这些信号的波长,即实现波长重用(wavelength reuse),提高了波长利用率.

在波长路由中由于节点内不含有波长转换器,因此,沿着一个路由在不同的链路上只能分配相同的波长给该路由.

虚波长路由(virtual wavelength routing) 网络理论术语.指在节点中包含有波长转换器.因此,沿着一个路由在不同的链路上可以分配不同的波长给该路由.和波长路由相比,虚波长路由可以更有效地提高波长利用率,降低网络阻塞率.

波长分配(wavelength assignment) 通信理论术语.指在网络中给已经建立的路由分配波长.在波长路由情况下,分配给一个路由的波长在不同的链路上必须相同,而在虚波长路由情况下,分配给一个路由的波长在不同的链路上可以不同.

路由和波长分配算法(routing and wavelength assignment algorithm) 一种网络算法.指在网络中给定了节点之间的连结请求后,首先在网络中为连结请求找到路由,然后给这些路由分配波长.实现路由和波长分配的算法就称为路由和波长分配算法.数学上已经证明路由和波长分配问题是NP-hard问题,因此,在大规模网络中,必须采用启发式算法进行求解.为了求解上的方便,路由和波长分配问题通常也分解为路由问题和波长分配问题,并进行分别求解.路由算法通常有最短光路路由、最小跳数路由和最小权重路由等.波长分配算法通常有首次命中算法、随机分配算法等.

保护(protection) 提高网络生存性的方式之一.指应用于环形网或格状(mesh)网络中的提高网络生存性的方式.保护方式是利用节点间预先(建网初期)分配好的容量,在发生故障时,通过保护倒换开关,直接将失效的通路用预先为其分配好的保护通路代替,其保护倒换速度较快.但是,由于网络资源很难共享,所以仍存在着资源浪费的现象.保护主要包括1+1保护、1∶1保护和1∶N保护等几种方式.

恢复(restoration) 提高网络生存性的方式之一.指应用于环形网或格状(mesh)网络中的提高网络生存性的方式.恢复方式是利用节点间的任何可用容量,在网络中为失效路由寻找一条新的替代路由.恢复包括:区段恢复与通道恢复两种重选路由方式,集中与分布两种控制机制.恢复方式的资源利用率较之保护方式要高,但是由于不是自动倒换,所以整个过程的耗时会有所增加.

多协议标签交换(multi-protocol label switching) 简称MPLS.一种将第二层交换和第三层路由结合起来的L2/L3集成技术.MPLS之所以称为

“多协议”是因为它不但可以支持多种第三层的网络协议，如 IPv4，IPv6，IPX，Appletalk，DECnet，CLNP 等，还可以同时兼容第二层上的多种链路层技术.

在网络的入口处，MPLS 首先根据某种特定的映射规则给数据包的前面加上一个固定长度的短标签，这种映射规则不但要考虑到数据包的目的地址信息，而且要考虑到有关服务质量信息. 在以后的网络转发过程中，支持 MPLS 的路由器（标签交换路由器 LSR：Label Switch Router）就只根据数据包所携带的标签进行转发或者交换. MPLS 最大的价值在于它能够在一个无连结的网络（例如 IP 网）中引入连结模式特性，从而有可能比较好地解决无连结型网络中的服务质量问题.

多协议波长交换（multi-protocol lambda switching） 简称 MPλS. 是多协议标签交换（参见“多协议标签交换”）技术为支持光网络内波长交换所做的扩展，MPλS 中光交叉连结节点 OXC（参见“光交叉连结”）相当于标签交换路由器，光信号的波长相当于 MPLS 中数据包的标签.

MPλS 技术是基于标签交换路由器和光交叉连结节点的相似之处提出来的. 从完成的基本功能看，它们都是将从任一个端口输入的数据从正确的端口输出，并且它们都是面向连结的. 对于标签交换路由器，输入和输出是通过如下的对应关系来确定的：<输入端口号 i，输入标签号 j；输出端口号 k，输出标签号 m>代表着一组完备的交换操作. 对于 OXC 有类似的关系，<输入光纤端口号 i，输入波长编号 j；输出光纤端口号 k，输出波长编号 m>.

广义多协议标签交换（generalized multi-protocol label switching） 简称 GMPLS. 是多协议标签交换（参见“多协议标签交换”）技术的推广. 推广后的 GMPLS 技术不但可以应用于包交换的网络，而且可以支持时间域、波长域和空间域的交换. 在多协议标签交换中，通过在数据包的前面增加一个固定长度的标签来标记数据包，而 GMPLS 对标签的概念进行了进一步扩充：对于时分复用的系统，如 SDH/SONET（参见“同步数字体系”），GMPLS 将信号所占的时隙作为其标签；对于波长交换的系统，如采用波长交换技术的光网络，选用光信号的波长作为其标签.

采用 GMPLS 技术可以将现有的几乎所有的网络及交换技术统一起来，形成一个统一的网络，这样的网络中可以包括因特网协议（IP）路由器、ATM（参见“异步传送模式”）交换机、同步数字体系（SDH）设备、光交叉连结（参见“光交叉连结”）等多种不同的硬件设施. GMPLS 技术则为它们提供一个统一的控制面，使得它们能够有机地集成在一起为用户服务.

数字包封（digital wrapper） 数字通信术语. 指朗讯（Lucent）公司提出的一种类似于 SONET（参见“同步数字体系”）的技术，主要是解决光通道的管理问题. 基本原理是提供一个数字包封容器，把各种业务容纳在其中，然后增加包含各种管理信息的信头，以及前向纠错（FEC）部分，形成一个传输单元. 这样一方面保证了系统的兼容性，可以容纳各种上层业务，另外也提供了非常丰富的管理字节，可以对光通道进行有效的管理和监控. 目前，它还没有成熟，处于发展中. 国际电信联盟的 G.709 草案中对其做了论述.

用户网络接口（user network interface） 简称 UNI. 网络理论术语. 指用户设备（包括 IP 路由器、ATM（参见“异步转移模式”）交换机、SDH（参见“同步数字体系”）交换机等）和智能光网络之间的接口，用户设备通过此接口向光网络提交获取、撤销和修改光带宽资源的申请，光网络对此申请请求进行应答. 用户设备和光网络之间基于客户-服务者关系模型. 该接口需要的规范内容主要有光通道速率、连结请求参数、命名和寻址方案、安全参数和响应时间等.

业务工程（traffic engineering） 亦称流量工程. 通信理论术语. 它是一种能将业务流映射到实际物理通路上，同时又可以自动优化网络资源以满足特定的服务性能要求，具有宏观调节和微观控制能力的网络工程技术. 传统的 IP 路由器进行路由处理时，未将带宽的可用性和业务特点等因素考虑进去，会因负载分布不均匀导致网络局部拥塞，这时就需要业务工程来解决此问题. 业务工程在数学上表现为一个规划问题，即在一定约束条件下求目标函数的最优值. 实践中比较常用的优化目标有以下几个：

1. 降低数据在传输和交换过程中引入的时延.

2. 均匀合理地分布网络负载，避免网络出现局部拥塞.

3. 合理分配备用资源，当网络出现故障时，业务可以比较容易地从路由到备用链路上，尽量减少故障对网络业务的破坏.

4. 满足网络管理员提出的一些策略性要求.

业务梳理（traffic grooming） 网络理论术语. 在以波长交换为基础、基于电路交换的波分复用技术（WDM）光网络中，波长是一种宝贵的资源，每个波长代表了很大的一块带宽资源. 而最终端的用户所申请使用的带宽一般是比较小的，需要经过电的逐级复用，将多个终端用户复用在一起然后再调制到某个光波长上. 如果在复用之前，没有经过仔细的预先筹划，不加区别的将用户信号复用在一起，则可能每经过一个节点，都需要进行光电变换并将一些

用户信号下路，这就增加了电处理设备的负担. 相反，假若能经过仔细的计算来选择复用方式就可以降低节点电子设备处理负担，充分发挥光网络的优势，这就是业务梳理所研究的内容.

信息理论

信息理论(information theory) 应用概率论、随机过程、数理统计、近世代数等数学方法来研究通信系统中信息传输的理论. 20 世纪 20 年代，奈奎斯特(Nyquist, H.)和哈特利(Hartley, R. V. I.)研究对各种电信系统进行比较的问题，提出了一些信息理论中的原理. 1948 年，仙农(Shannon, C. E.)发表了经典著作《通信的数学理论》，奠定了信息理论的数学基础，使信息理论成为重要的数学分支之一，称为仙农信息论.

一个信息传输系统的信息发生源，称为信源；传输信息的通道，称为信道；最后信息被传到接收端，称为信宿. 在信源和信道之间要经过编码，在信道与信宿之间需要译码. 有时由于系统中存在噪声，所接收到的最终信号是有错误的，称这种信道为有噪信道，否则称为无噪信道. 仙农信息论的基本内容就是研究信源、信道、编码、译码、信宿等问题. 信源的关键问题是如何定量表示它含有多少信息量；信宿的问题是能收到或提取多少信息量；信道的问题是它最多能传送多少信息(即信道容量)；编码、译码问题是如何使信源的信息被充分表达，信道容量被充分利用，以及如何安全可靠地传送信息. 与信源有关的另一类问题是失真问题，它研究在允许失真时必须传送的信息量.

随着信息科学技术的发展，以物质和能量为中心的传统科学观念已开始转化为以信息、物质和能量作为自然界和人类社会的三大基本资源的新的科学观念. 从这一观念出发，信息问题的研究已远远超出了仙农信息论的范畴，形成了信息科学.

信息(information) 信息科学中最基本、最重要的概念之一. 它的定义可以分为两个层次：

1. 纯客观的层次(即本体论的层次). 信息定义为"事物运动的状态和方式". 在这里，"事物"是泛指一切范畴的事物；"运动"是指任何意义上的运动；"状态"是反映相对稳定的一面；"方式"则反映运动变化的一面.

2. 使用的层次(即认识论层次). 信息定义为"关于事物运动的状态和方式的广义知识". 所谓广义知识，既包括一般系统化了的知识，也包括那些表述事物运动和方式的源知识. 信息、知识和情报是不能等同的. 知识仅是一种系统化的信息，是信息中被人类总结出的精华部分. 情报是知识中动态的、按特定方式获取和传递、为特定目的服务、并带有机密性和时间性的那一部分. 知识、情报、信息三者关系可以用三个圆来表示(见图 1). 它们之间的转化关系如图 2.

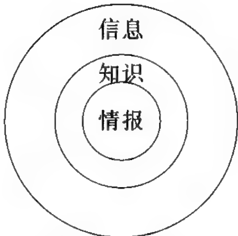


图1 情报、知识、信息关系图

从认识论观点出发，信息本身可以分为三个基本层次，即语法信息、语义信息和语用信息. 它们之间的关系如图 3 所示. 语法信息是信息的基本形态，语义信息是信息的含义，语用信息是信息的效用. 三者不是彼此独立的，而是密切相关、互相重叠的. 了解掌握信息的三个基本层次，可以加深对信息本质的认识，是正确运用信息的重要条件. 人和事物打交道，总是以信息作为中间媒介，而且总是首先得到语法信息，进而得到语义信息和语用信息. 然后把所得到的信息同目标信息进行比较、分析，做出相应的决策，再对事物的运动发展进行控制、调整，以期按自己的目标向前推进. 没有这样的过程，人们就无法进行有成效的实践活动. 人与三层次信息关系如图 4 所示.

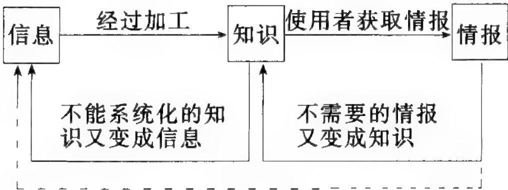


图2 信息、知识和情报转换示意图

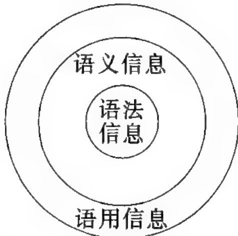


图3 语法信息、语义信息、语用信息关系图

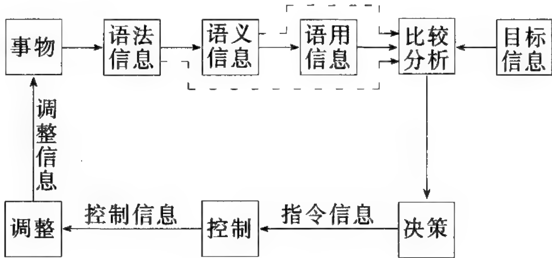


图4 人和信息的一般关系

信息具有下述特征：

1. 信息源于物质，但又不是物质本身，它可以脱离源物质而相对地独立存在.

2. 信息也源于精神，但又不限于精神领域.

3. 信息和能量息息相关，但又与能量有质的区别. 能量提供动力，信息提供的是知识和智慧.

4. 信息具有知识本性，但它又比知识含义更广泛.

5. 信息可以被感知、提取、传输、处理、存贮和使用等.

6. 信息可以被众多的用户共同分享.

7. 语法信息可以在传递和处理过程中永不增值.

信息特征(feature of information)　见“信息”.

先验概率(prior probability)　信息理论基本概念之一. 在一个通信系统中，在收到任何消息之前，接收端所了解到的某消息发送的概率，称为先验概率. 通过该系统所获得的信息量，随先验概率的增加而减少. 即定义为

$$\text{接收信息量} = \lg\left(\frac{\text{后验概率}}{\text{先验概率}}\right).$$

语法信息(syntactic information)　信息的最基本层次. 它只涉及事物运动的结构，即只考虑状态和状态之间的关系，正像语法只考虑词和词之间的关系那样. 它涉及符号的数目、它们持续的时间、信源的统计性质、编码方式等. 在这个层次上，不考虑这些信息的含义和效用. 语法信息比较抽象、单纯，相对地说比较容易处理. 因此，语法信息的理论(包括描述和数值度量理论)比较成熟. 仙农(Shannon, C. E.)所研究的就属于这个范围.

语义信息(semantic information)　信息的含义层次. 它不仅要考虑状态和状态之间的关系，还要考虑它们所具有的含义，即通过符号所表达的含义，也即信息包含的内容以及信息在逻辑上的真实性；但它还没有考虑信息使用者的个人主观因素. 由于引入了信息的含义，使该理论更加有用，也更加复杂. 因而，语义信息理论目前还没有形成自己的体系，处于探索阶段.

语用信息(pragmatic information)　信息的最高层次. 它必须以语法、语义信息为基础，即它不仅要考虑状态和状态之间关系以及它们的含义，还要进一步考察这种关系及含义对于信息使用者的效用和价值. 它比语义信息更依赖于接收者，而且与时间关系密切. 一个过时的信息，将会变得毫无价值，从中得不到有用的信息. 可见，对于语用信息，不仅包含了语义信息这种复杂因素，而且还要包含效用这种带主观意义的因素，因而使该理论更加符合实际，但也更加复杂.

数据(data)　信息科学中最基本、最重要的概念之一. 数据是信息的表现形态之一. 其他的表现形态还有文字、话音、图形、图象等.

概率信息(probabilistic information)　随机试验(或随机事件)提供的信息. 可以利用概率空间的方法来表示这类信息. 概率空间本身就表示了事物运动方式是随机性的，而概率空间各元素就表示了该事物可能的运动状态. 假定 x 为任意随机试验，它的可能状态集合为 X，这个集合的先验概率测度为 P，后验概率测度为 P^*，则概率空间的变换为

$$\begin{Bmatrix} X \\ P \end{Bmatrix} \Rightarrow \begin{Bmatrix} X \\ P^* \end{Bmatrix},$$

其中 $\begin{Bmatrix} X \\ P \end{Bmatrix}$ 表示观察者的先验信息，这就是先验概率空间. $\begin{Bmatrix} X \\ P^* \end{Bmatrix}$ 表示后验概率空间. X 可以是离散型的，也可以是连续型的. 上式表示利用后验概率空间来描述实际信息，利用先验概率空间来表示先验信息，利用这两个概率空间的变换来表示实得信息，这就完成了概率信息的表示.

确定信息(deterministic information)　具有确定规律事件提供的信息. 如某事物(或一个系统)的运动状态和运动方式可以由微分方程来描述. 比如，某系统的行为由一个 n 阶线性常微分方程给出：

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y^{(1)} + a_n y = u.$$

若没有外来随机干扰，只要给定初始条件 $y(0)$, $y^{(1)}(0), \cdots, y^{(n-1)}(0)$ 和 $t \geqslant 0$ 时的输出 u，就完全可以确定这一系统在未来任意时刻 t 的行为. 在实际工作中，可以用状态空间法来表示这种系统的行为. 而状态空间法的状态方程就是直接由微分方程导出来的. 例如，取 $y, y^{(1)}, y^{(2)}, \cdots, y^{(n-1)}$ 为 n 个状态参量，并令

$$\begin{cases} x_1 = y, \\ x_2 = y^{(1)}, \\ x_3 = y^{(2)}, \\ \cdots\cdots\cdots\cdots \\ x_n = y^{(n-1)}. \end{cases}$$

于是，上述 n 阶线性微分方程就可以写成

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3, \\ \dot{x}_3 = x_4, \\ \cdots\cdots\cdots \\ \dot{x}_{n-1} = x_n, \\ \dot{x}_n = -a_n x_1 - a_{n-1} x_2 - \cdots - a_1 x_n + u. \end{cases}$$

这就是状态方程，写成矩阵形式为

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_{n-1} \\ \dot{x}_n \end{Bmatrix} = \begin{Bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{Bmatrix} \cdot \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{Bmatrix} + \begin{Bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ u \end{Bmatrix}.$$

模糊信息(fuzzy information)　模糊事件提供的信息. 设有一模糊集合 X，它有 N 个元素：$x_1, x_2, \cdots, x_N$，各个元素的隶属度分别为 $f_1, f_2, \cdots, f_N$. 这个“运动状态和方式”就可以用相应的隶属度空间来表示. 如果经过某种处理(如使模糊的照片变为清晰的照片)，各个元素的隶属度分别变为 f_i^* $(i=1,2,\cdots,N)$，那么，与这个过程相应的信息就可以由相应的隶属度空间变换来表示，即

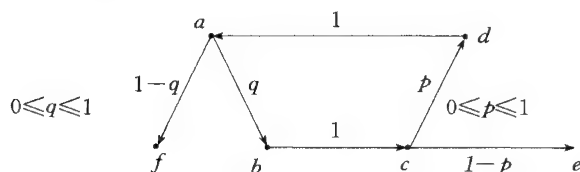
$$\begin{Bmatrix} X \\ F \end{Bmatrix} \Rightarrow \begin{Bmatrix} X \\ F^* \end{Bmatrix},$$

$$\{X\} = \{x_1, x_2, \cdots, x_N\},$$

$$\{F\} = \{f_1, f_2, \cdots, f_N\},$$

$$\{F^*\} = \{f_1^*, f_2^*, \cdots, f_N^*\}.$$

混合信息(mixed information)　具有多种信息类型的信息. 某一事件的运动方式在不同时段分别具有概率、模糊和确定的特性，由该事件提供的信息称为混合信息. 可以采用图论的方法表示这种信息，图的顶点用来表示状态，加权边用来表示运动方式. 例如，下图就是一个描述一年生植物的生活行为的状态转移图. 在图中，各个顶点表示植物生活状态，连线表示状态转移途径，箭头表示转移方向，连线旁边数字(称为权，这样的连线称加权边)表示从一个状态向另一个状态转移的概率. 下图表示：种子 a 生长为植物 b 的概率为 q，以概率 $1-q$ 成为死亡种子 f；植物 b 肯定开花为植物 c，开花植物 c 以概率 p 受粉成为一种受粉植物 d，它必定结出种子 a. 开花植物 c 以概率 $1-p$ 成为未受粉植物 e.

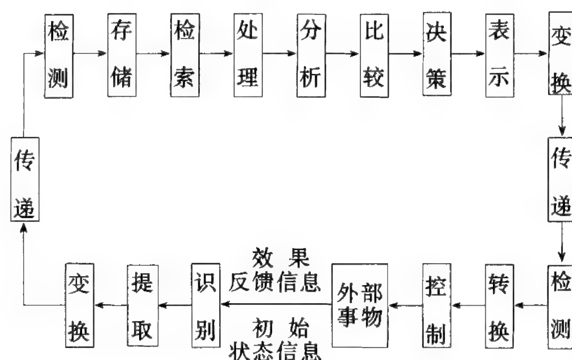


后验概率(posterior probability)　信息理论的基本概念之一. 在一个通信系统中，在收到某个消息之后，接收端所了解到的该消息发送的概率称为后验概率. 通信所得到的信息量，随先验概率的增加而减少，随后验概率的增加而增加. 通常定义为

$$\text{接收信息量} = \lg\left(\frac{\text{后验概率}}{\text{先验概率}}\right).$$

信息科学(information science)　研究信息现象及其规律的一门学科. 是以信息为主要研究对象，以信息的运动规律和利用信息的原理为主要研究内容，以信息科学方法论为主要研究方法，以扩展人的信息功能(特别是智力功能)为主要研究目标的一门新兴学科.

由于信息科学的崛起和发展，以物质和能量为中心的传统自然科学体系才能发展为信息、物质和能量三位一体的现代自然科学完整体系. 信息科学的主体结构包含信息论、控制论和系统论三部分. 人工智能理论实质上是这三者的综合运用. 下图为信息科学基本问题间的关系.

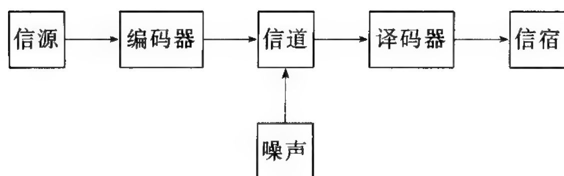


信息科学方法论(methodology of information science)　研究信息科学基本方法的理论. 信息科学方法论的具体内容包括三个方面，这就是信息分析和综合方法，行为功能模拟准则，以及系统整体优化准则. 简单地说，就是信息方法、功能模拟准则和系统整体优化准则.

信息方法的内容有两个基本部分. 第一是信息分析方法，即从信息的观点出发，去认识分析复杂事物工作机制的信息过程. 第二是信息综合方法，即从信息的观点出发，去实现复杂事物功能的方法. 功能模拟准则即在利用信息分析方法理解自然系统工作机制的基础上，利用信息综合方法来综合相应的人工系统时，应对它的自然系统进行功能模拟，即以行为的相似为基础，从功能上模拟原型的信息过程以及它与环境相互作用的方式. 所谓系统整体优化准则，是信息方法的一个实施性准则，它告诉人们要寻求系统整体性能最优，而不是局部要素性能的最优.

信息技术(information technique)　信息科学研究的内容之一. 指可以扩展人的信息器官功能的技术. 信息技术包含内容甚多，但就其主体来说，它的最重要、最典型的分支是传感技术、通信技术、计算机技术和控制技术. 这是一切技术中最基本也是最精华的部分. 传感技术主要包括信息的识别、检测、提取、变换等技术，它是人的信息感知功能的扩展和延伸. 通信技术包括信息的变换、传递、控制、调节等技术，它是人的信息传递系统(神经系统)功能的扩展和延伸. 计算机技术包括信息的存储、检索、处理、分析、控制等技术，它是人的信息处理器官(大脑)功能的延伸.

仙农通信系统模型(Shannon communication system model)　一种基本的通信系统模型. 仙农通信系统由信源、编码器、信道、译码器和信宿等组成. 该系统的模型如图所示.



信源是信息的发出者.由于信源的信息形式不同,就要求有不同的通信系统来与之相适应,从而形成了多种多样的通信系统.如:信息形式为语言的信源,则有电话通信系统;信息形式为文字的信源,则有电报及传真通信系统;信息形式为图象的信源,则有图象通信系统;信息形式为计算机数据的信源,则有计算机通信系统等.总之,信源形式的不断增多,通信系统的形式也越来越多样化.

信 源(information source) 信息科学中最基本、最重要的概念之一.是信息的产生与发出者,它可通过消息来表达.信源可根据消息出现的形式分为两大类型:离散信源和连续信源.在通信系统中,收信者在未收到消息以前,对信源发出什么消息是不确定的,所以可用随机变量或随机变量序列来描述信源输出的消息.信源的数学模型是一个在信源符号集中取值的随机变量序列或随机过程.

离散信源(discrete information source) 信源的一种类型.有的信源输出的消息常常是以一个个符号的形式出现,且符号的取值是有限的或可数无穷的,则称该信源为离散信源.所以,可以用离散型随机变量来描述这些消息.它的数学模型就是在信源符号集 $A=\{a_1,a_2,\cdots,a_q\}$ 中取值的随机变量 X 或随机变量序列 $X_1,X_2,\cdots,X_i,\cdots$,其中 X_i 表示信源在时刻 $t=i$ 输出的消息.它取值为符号 a_j 的概率为

$$P_r\{X_i=a_j\}=P(a_j)\geqslant 0\ (j=1,2,\cdots,q),$$

且满足

$$\sum_{j=1}^{q}P(a_j)=1.$$

离散平稳信源(discrete stationary source) 离散信源的一种类型.若在所有时刻 $t=i$ $(i=1,2,\cdots)$,$P_r\{X_i=a_j\}=P(a_j)$,则符号序列是一维平稳的,这表示任意两个不同时刻信源发出的符号的概率分布完全相同,即

$$P_r\{X_i=a_1\}=P_r\{X_j=a_1\}=P(a_1),$$
$$P_r\{X_i=a_2\}=P_r\{X_j=a_2\}=P(a_2),$$
$$\cdots\cdots$$
$$P_r\{X_i=a_q\}=P_r\{X_j=a_q\}=P(a_q).$$

若各维联合概率分布均与时间起点无关,即当 $t=i$,$t=j$ $(i,j$ 为任意整数,且 $i\neq j)$时有

$$P(x_i)=P(x_j),$$
$$P(x_i,x_{i+1})=P(x_j,x_{j+1}),$$
$$\cdots\cdots$$
$$P(x_i,x_{i+1},\cdots,x_{i+N})=P(x_j,x_{j+1},\cdots,x_{j+N}),$$

其中

$$P(x_i),\ P(x_i,x_{i+1}),\ \cdots,\ P(x_i,x_{i+1},\cdots,x_{i+N})$$

分别为

$$P_r\{X_i=x_i\},\ P_r\{X_i=x_i,X_{i+1}=x_{i+1}\},\cdots,$$
$$P_r\{X_i=x_i,X_{i+1}=x_{i+1},\cdots,X_{i+N}=x_{i+N}\}$$

的简单记法.则信源是完全平稳的,信源发出的符号序列也是完全平稳的,称完全平稳的信源为平稳信源.

对于离散平稳信源,当熵 $H(X_1)<\infty$时,具有以下性质:

1. 条件熵 $H(X_N|X_1,X_2,\cdots,X_{N-1})$随 N 的增加是非递增的.

2. $H_N=\frac{1}{N}H(X_1,X_2,\cdots,X_{N-1})$
$\geqslant H(X_N|X_1,X_2,\cdots,X_{N-1})$.

3. H_N 随 N 的增加是非递增的,因而是有界的.

4. $\overline{H}=\lim\limits_{N\to\infty}\frac{1}{N}H(X_1,X_2,\cdots,X_N)$存在,且
$\overline{H}=\lim\limits_{N\to\infty}H(X_N|X_1,X_2,\cdots,X_{N-1})$,

这里 $\overline{H}$ 称为平稳信源的熵率.

离散平稳信源的熵率(entropy rate of discrete stationary source) 见"离散平稳信源".

离散无记忆信源(discrete memoryless source) 离散信源的一种类型.若表示信源输出消息的随机变量序列是一个相互独立同分布的随机变量序列,在此情况下,信源先后发出的一个个符号是统计独立同分布的,即对任意正整数 N,其 N 维联合概率分布满足

$$P(x_1,x_2,\cdots,x_N)=\prod_{i=1}^{N}P(X_i),$$

换句话说,随机变量序列中的任意 N 维随机矢量的联合概率分布可用随机矢量中单个随机变量的概率分布的乘积来表示.这种信源称为离散无记忆信源.离散无记忆信源可用它输出的一个随机变量 X 来表示.

离散有记忆信源(discrete memorable source) 离散信源的一种类型.在一般情况下,信源先后发出的符号之间是互相关联的.例如,在中文字母组成的中文消息中,前后文字的出现是有关联的,故在 N 维随机矢量的联合概率分布中,引入条件概率分布来描述它们之间的关联.称这种信源为离散有记忆信源.实际上信源发出的符号往往只与前若干个符号的关联较强,而与更前面的符号关联较弱.可以用"记忆长度"来描述这种关联性.

保真度准则(fidelity criterion) 信息理论的基本概念之一.指信源压缩过程中的保真度标准.假定

某一信源的允许平均失真度为 D^*，则信源压缩后单个符号的平均失真度 $\overline{D}$ 满足 $\overline{D}\leqslant D^*$，称此不等式为保真度准则. 长为 K 的信源符号序列的保真度准则为

$$\overline{D}(K)\leqslant KD^*.$$

当信源固定，单个符号失真度固定时，选择某一信道满足 $\overline{D}\leqslant D^*$，称这种信道为 D^* 失真许可试验信道. 由于满足这一条件的信道很多，可将它们组成一个集合，用 B_{D^*} 表示，即

$$B_{D^*}=\{P(v_j|u_i),\overline{D}\leqslant D^*\},$$

$$B_{D^*(K)}=\{P(\beta_j|\alpha_i),\overline{D}(K)\leqslant KD^*\}.$$

信息率失真函数(information rate distortional function) 信息理论的基本概念之一. 在信源给定时，总希望在满足一定失真的情况下信息传输率 R 尽可能小. 即满足 $\overline{D}\leqslant D^*$ 时，寻找信息传输率 R 的下限值. 从接收端看，即在 $\overline{D}\leqslant D^*$ 情况下，寻找平均互信息量 $I(U,V)$ 的最小值. 设在 D^* 失真许可试验信道中某一信道 $P(v_j|u_i)$ 在 $\overline{D}\leqslant D^*$ 条件下有

$$R(D^*)=\min\{I(U,V),P(v_j|u_i)\in B_{D^*}\},$$

称该函数为信息率失真函数. 对于长度为 K 的信源符号序列，信息率失真函数为

$$R_K(D^*)=\min\{I(U,V_K),P(\beta_j|\alpha_i)\in B_{D^*(K)}\}.$$

$R(D^*)$ 反映了在满足 $\overline{D}\leqslant D^*$ 情况下信源可压缩的程度.

信源率失真函数(source rate distortional function) 信息理论的基本概念之一. 信源率失真函数定义为

$$R(D^*)=\min\{I(U,V),P(v_j|u_i),\overline{D}\leqslant D^*\}.$$

信源率失真函数的性质有：

1. $R(D^*)$ 的定义域为 $(0,D^*_{\max})$.
2. $R(D^*)$ 是允许失真度 D^* 的下凸函数.
3. $R(D^*)$ 函数具有单调递减性.
4. $R(D^*)$ 函数具有连续性.

信源剩余度(redundancy of source) 信息理论的基本概念之一. 信源的剩余度等于 1 减去熵的相对率，即

$$r=1-\eta=1-\frac{\overline{H}}{\lg q},$$

其中 $\overline{H}$ 为信源的熵率，q 为信源符号数. 由此式可见，信源符号之间依赖关系越大，$\overline{H}$ 越小，信源的剩余度就越大. 从提高传输信息效率的观点出发，总是希望减少或去掉信源的剩余度. 但从减少干扰的影响观点出发，信源的剩余度越大，消息抗干扰能力越强.

自信息量(self-information) 信息理论的基本概念之一. 即随机事件所固有的信息量. 一随机事件 A，它的概率为 $P(A)$. $P(A)$ 越大，事件 A 发生的可能性越大，不确定性越小；$P(A)$ 越小，事件 A 发生的可能性越小，不确定性越大. 用 $P(A)$ 的对数反映事件 A 的不确定性程度，即

$$I_A=\lg\frac{1}{P(A)}=-\lg P(A),$$

I_A 称为随机事件 A 的自信息量.

若对数以 2 为底，则 I_A 的单位是比特；若对数以 10 为底，则 I_A 的单位是迪西特；若对数以 e 为底，则 I_A 的单位是奈特.

互信息量(mutual information) 信息理论的基本概念之一. 即由一个事件反映另一个事件的信息量. 互信息量 $I(x;y)$ 是代表收到某消息 y 后所获得的关于某事件 x 的信息量，即

$$I(x;y)=\lg\frac{P(x|y)}{P(x)}=\lg\frac{P(x,y)}{P(x)P(y)}=\lg\frac{P(y|x)}{P(y)}.$$

互信息量可取正值，也可取负值. 若 $I(x;y)$ 为负，说明收信者在未收到消息 y 以前对消息 x 是否出现的猜测疑难程度较小. 但由于噪声的存在，接收到消息 y 后，反而使收信者对消息 x 是否出现的猜测难度增加了. 也就是说收信者接收到消息 y 后对 x 出现的不确定性反而增加，所以获得的信息量为负.

平均互信息量(average mutual information) 信息理论的基本概念之一. 已知 $H(X)$ 代表接收到随机消息 Y 以前关于随机消息 X 的平均不确定性，而 $H(X|Y)$ 代表接收到随机消息 Y 后关于随机消息 X 的平均不确定性. 可见，通过信道传输消息减少了一些不确定性，消耗了一定的信息量. 故定义

$$I(X;Y)=H(X)-H(X|Y),$$

称 $I(X;Y)$ 为 X 和 Y 之间的平均互信息量.

$$I(X;Y)=\sum_{x,y}P(x,y)\lg\frac{P(x|y)}{P(x)}=\sum_{x,y}P(x,y)\lg\frac{P(x,y)}{P(x)P(y)}=\sum_{x,y}P(x,y)\lg\frac{P(y|x)}{P(y)}\quad(x\in X,y\in Y),$$

其中 $x\in A,y\in B,A,B$ 分别为信道的输入和输出符号集. 从平均互信息量的定义中，可以进一步理解熵只是平均不确定性的描述，而不确定性的消除(熵差)才等于接收端所获得的信息量. 平均互信息量与信源和信道的关系如下：

1. 平均互信息量 $I(X;Y)$ 是输入信源的概率分布 $P(x)$ 的上凸函数. 即当信源发出的消息通过给定信道传输时，选择不同的信源(其概率分布不同)，在信道输出端接收到相应消息后获得的信息量是不同的，$I(X;Y)$ 与 $P(x)$ 的关系为上凸函数，而且对于每

一个固定信道，一定存在一种信源，其概率分布为 $P(x)$，使输出端获得的平均信息量为最大（因上凸函数存在极大值）.

2. 平均互信息量 $I(X;Y)$ 是信道传递概率 $P(y|x)$ 的下凸函数. 即当信源给定后，选择不同的信道来传输同一信源发出的消息时，在输出端获得关于信源的信息量是不同的，信道输出端获得关于信源的信息量是信道传递概率的下凸函数，且对每一种信源都存在一种最差的信道，此信道干扰（噪声）最大，而输出端获得的信息量最小.

平均互信息量 $I(X;Y)$ 的基本特性有：

1. 平均互信息量的非负性. $I(X;Y)\geqslant 0$，当 X 和 Y 统计独立时，等式成立.

2. 平均互信息量的极值性.

$$I(X;Y)\leqslant \min\{H(X),H(Y)\}.$$

3. 平均互信息量的交互性（对称性）. 即

$$I(X;Y)=I(Y;X).$$

4. 平均互信息量与各类熵的关系.

$$\begin{aligned}I(X;Y)&=H(X)-H(X|Y)\\&=H(Y)-H(Y|X)\\&=H(X)+H(Y)-H(XY).\end{aligned}$$

5. 平均互信息量 $I(X;Y)$ 对分布函数具凸性.

马尔可夫信源（Markov source） 一种离散有记忆信源. 即任何时候信源发出符号的概率只与前面已经发出的数个符号有关，而与更前面发出的符号无关. 若发出符号的概率只与前面发出的 m 个符号有关时，则称此信源为 m 阶马尔可夫信源. 它的数学模型由信源符号集 A 上的一组条件概率确定：

$$A=[a_1,a_2,\cdots,a_q],$$

$$P(a_{k_{m+1}}|a_{k_1}a_{k_2}\cdots a_{k_m})\geqslant 0$$

$$(k_1,k_2,\cdots,k_m,k_{m+1}=1,2,\cdots,q),$$

并满足

$$\sum_{k_{m+1}=1}^{q}P(a_{k_{m+1}}|a_{k_1}a_{k_2}\cdots a_{k_m})=1.$$

失真函数（distortional function） 信息理论的基本概念之一. 它表示符号在信道传输过程中所引起的失真的大小. 设离散无记忆信源符号集

$$U=[u_1,u_2,\cdots,u_r],$$

其概率分布为

$$P(u)=[P(u_1),P(u_2),\cdots,P(u_r)].$$

信源符号通过信道传输到接收端，接收端符号集为

$$V=[v_1,v_2,\cdots,v_s].$$

对于每一对 $[u,v]$，定义一个非负函数

$$d(u_i,v_j)\geqslant 0\quad (i=1,2,\cdots,r,\ j=1,2,\cdots,s),$$

称该函数为失真函数，又称单个符号的失真度. 用它来测度信源发出一个符号 u_i 后在接收端收到一个符号 v_j 时所引起的误差，或称失真. 失真函数 $d(u_i,v_j)$ 可用矩阵形式表示为

$$[D]=\begin{bmatrix}d(u_1,v_1) & d(u_1,v_2) & \cdots & d(u_1,v_s)\\ d(u_2,v_1) & d(u_2,v_2) & \cdots & d(u_2,v_s)\\ \vdots & \vdots & & \vdots\\ d(u_r,v_1) & d(u_r,v_2) & \cdots & d(u_r,v_s)\end{bmatrix},$$

称该矩阵为失真矩阵，它是 $r\times s$ 阶矩阵.

失真矩阵（distortion matrix） 见“失真函数”.

信源平均失真度（average distortion of source） 信息理论的基本概念之一. 在离散情况下，信源符号集 $U=[u_1,u_2,\cdots,u_r]$，其概率分布为

$$P(u)=[P(u_1),P(u_2),\cdots,P(u_r)],$$

接收端符号集为 $V=[v_1,v_2,\cdots,v_s]$. 若已知信道的传递概率为 $P(v|u)$ 时，则定义信源平均失真度为

$$\begin{aligned}\overline{D}&=\sum_{u,v}P(u,v)d(u,v)\\&=\sum_{i=1}^{r}\sum_{j=1}^{s}P(u_i)P(v_j|u_i)d(u_i,v_j).\end{aligned}$$

此式说明，信源平均失真度是对信源和信道的统计平均值，它描述某一信源的概率分布 $P(u_i)$ 在给定传递概率为 $P(v|u)$ 的某一信道传输下的失真大小.

熵（entropy） 信息理论的基本概念之一. 它是平均信息量的量度. 设 X 是取有限值的离散随机变量，其值集为

$$A=\{x_1,x_2,\cdots,x_N\},\quad P(x_i)=P_r\{X=x_i\},$$

记为 $\{A,P\}$. 事件 $\{x=x_i\}$ 的自信息量 $-\lg P(x_i)$ 的平均值为

$$H(X)=-\sum_{i=1}^{N}P(x_i)\lg P(x_i),$$

称 $H(X)$ 为 $\{A,P\}$ 的熵. 设 X,Y 为两个离散随机变量，即

$$P(x_i,y_j)=P_r\{X=x_i,Y=y_j\},$$

$$P(y_j|x_i)=P_r\{Y=y_j|X=x_i\},$$

则条件熵定义为

$$H(Y|X)=-\sum_{i}\sum_{j}P(x_i,y_j)\lg[P(y_j|x_i)].$$

平均互信息量为

$$I(X;Y)=H(Y)-H(Y|X).$$

信源联合熵（joint entropy of source） 信息理论的基本概念之一. 将信源 X 输出的随机序列每两个符号分成一组，并不考虑组与组之间的统计. 这时，可等效成一个新的信源 X_1X_2，它们的联合概率空间为

$$\{X_1X_2,P(X_1X_2)\}=\{a_1a_1,a_1a_2,\cdots,a_{q-1}a_q,a_qa_q;\ P(a_ia_j)=P(a_i)P(a_i|a_j)\}.$$

根据信息熵定义，可得

$$H(X_1X_2)=-\sum_{i=1}^{q}\sum_{j=1}^{q}P(a_ia_j)\lg P(a_ia_j).$$

称 $H(X_1X_2)$ 为信源 X 的联合熵.

信源条件熵(conditional entropy of source) 信息理论的基本概念之一.将信源 X 输出的随机序列每两个符号分成一组,并不考虑组与组之间的统计.这时,可等效成一个新的信源 X_1X_2,其中 X_1(或 X_2)可取 $A=\{a_1,a_2,\cdots,a_q\}$ 中任一个,均存在一个熵 $H(X_2|X_1=a_i)$,对所有 a_i 的可能值进行统计平均,就可求得.当前面一个符号已知时,再输出后面一个符号的总的熵

$$H(X_2|X_1)=-\sum_{i=1}^{q}\sum_{j=1}^{q}P(a_ia_j)\lg P(a_j|a_i),$$

称 $H(X_2|X_1)$ 为信源 X 的条件熵.

熵的意义(meaning of entropy) 对信息熵的解释.设离散信源 X 的概率空间为

$$\{A,P(x)\}=\{a_1,a_2,\cdots,a_q;\ P(a_1),P(a_2),\cdots,P(a_q)\},$$

若其中某一事件 a_i 的自信息量为

$$I(a_i)=-\lg P(a_i),$$

则信源 X 中所有事件的自信息量的平均值为

$$H(X)=-\sum_{i=1}^{q}P(a_i)\lg P(a_i),$$

称 $H(X)$ 为信源的信息熵(亦称信源 X 的平均信息量).信息熵的物理含义为:

1.信息熵 $H(X)$ 表示信源输出消息后所提供的平均信息量.

2.信息熵 $H(X)$ 表示信源输出消息前,信源的平均不确定性.

3.信息熵 $H(X)$ 反映了变量 X 的随机性.

熵的代数性质(algebraic property of entropy) 熵函数的特性之一.因为信息熵 $H(X)$ 是概率 $P_1=P(a_1),P_2=P(a_2),\cdots,P_q=P(a_q)$ 的函数,故可表示为

$$H(P)=H(P_1,P_2,\cdots,P_q)=-\sum_{i=1}^{q}P_i\lg P_i,$$

即熵 H 是 $P_1,P_2,\cdots,P_q$ 的 q 元函数.$H(P)$ 有如下性质:

1.对称性.当变量 $P_1,P_2,\cdots,P_q$ 的顺序任意重排时,熵的值不变.特别有

$$H(P_1,P_2,\cdots,P_q)=H(P_2,P_3,\cdots,P_q,P_1)=H(P_q,P_1,\cdots,P_{q-1}).$$

2.确定性.当某变量 $P_i=1$ 时,则 $H(P)=0$.

3.非负性.$H(x)\geqslant 0$(对离散信源).

4.扩展性.

$$\lim_{\varepsilon\to 0}H_{q+1}(P_1,P_2,\cdots,P_q-\varepsilon,\varepsilon)=H(P_1,P_2,\cdots,P_q).$$

5.可加性.当 X,Y 统计独立时,有

$$H(X,Y)=H(X)+H(Y|X).$$

熵的解析性质(analytic property of entropy) 熵函数的特性之一.仙农熵函数

$$H_n(P_1,P_2,\cdots,P_q)=\sum -P_i\lg P_i$$

具有下述解析性质.首先要注意不等式

$$\lg x\leqslant x-1\quad(x>0).$$

1.对任何概率分布 $P_1,P_2,\cdots,P_q$,

$$H_n(P_1,P_2,\cdots,P_q)\leqslant H_n\left(\frac{1}{n},\frac{1}{n},\cdots,\frac{1}{n}\right)=\lg n,$$

尤其当 $n=2$ 时,可写成 $H(P)=H_2(P,1-P)$,如图,它是上凸函数,且在 $P=1/2$ 时有极大值.

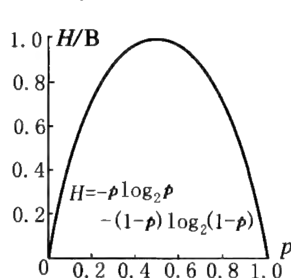
2.对于任意的 $i=1,2,\cdots,n$,若 $(P_{1i},P_{2i},\cdots,P_{mi})$ 为 m 维概率向量,$P=(P_1,P_2,\cdots,P_n)$ 为 n 维概率向量,则一般地有

$$\sum_{i=1}^{n}P_iH_m(P_{1i},P_{2i},\cdots,P_{mi})\leqslant H_m\left(\sum_{i=1}^{n}P_iP_{1i},\sum_{i=1}^{n}P_iP_{2i},\cdots,\sum_{i=1}^{n}P_iP_{mi}\right).$$

3.对于任意的 n 维概率向量 $P=(P_1,P_2,\cdots,P_n)$ 和 $q=(q_1,q_2,\cdots,q_n)$,

$$H_n(P_1,P_2,\cdots,P_n)=\sum_{i=1}^{n}(-P_i)\lg P_i\leqslant\sum_{i=1}^{n}(-P_i)\lg q_i,$$

等式仅在 $P=q$ 时成立.这是编码理论中常用的基本不等式之一,有时亦称仙农引理.



4. $H(X,Y)\leqslant H(X)+H(Y)$.

等号成立当且仅当 X,Y 统计独立.

最大离散熵定理(maximal discrete entropy theorem) 离散信源的熵达到最大值的一个重要定理.该定理断言:对于具有 q 个符号的离散信源,当 q 个信源符号出现的可能性相等时,信源的熵达到最大值,即

$$H(P_1,P_2,\cdots,P_q)\leqslant H\left(\frac{1}{q},\frac{1}{q},\cdots,\frac{1}{q}\right)=\lg q.$$

称信源中的事件为等概率事件.即等概率分布的信源其平均不确定性最大.

熵的强可加性(strong additivity of entropy) 熵的一种性质.联合熵等于前一个符号出现的熵加上前一个符号已知时后一个符号出现的条件熵,即

$$H(X_1X_2)=H(X_1)+H(X_2|X_1),$$

称该性质为熵的强可加性.当前后符号无依赖关系时,该性质即变为熵的可加性,即

$$H(X,Y)=H(X)+H(Y).$$

熵的相对率(relativity of entropy)　关于熵的一个比值. 指一个信源的熵率与具有同样符号集的最大熵的比值,即

$$\eta = \frac{\overline{H}}{\lg q},$$

式中 $\overline{H}$ 是信源的熵率,$\lg q$ 是信源的最大熵,q 为信源的符号数.

离散信道(discrete channel)　信道的一种类型. 传输离散信源的信道称为离散信道. 离散信道的数学模型如下图所示. 图中信道为 $P(Y|X)$,



$$X=(X_1,X_2,\cdots,X_N),\quad Y=(Y_1,Y_2,\cdots,Y_N),$$

符号集 $A=[a_1,a_2,\cdots,a_r]$和 $B=[b_1,b_2,\cdots,b_s]$分别为信道的输入和输出信号集,亦称信道的入口和出口符号集,图中 X 和 Y 为随机矢量. 每个随机变量 X_i 和 Y_i 分别取值于符号集 $A=[a_1,a_2,\cdots,a_r]$和 $B=[b_1,b_2,\cdots,b_s]$. $P(Y|X)$为条件概率,它描述了输入信号和输出信号之间统计依赖关系,反映了信道的特性. 根据条件概率的不同,离散信道可分成三种情况:

1. 无干扰(无噪)信道.
2. 有干扰无记忆信道.
3. 有干扰有记忆信道.

离散信道有如下几种概率:

1. 输入和输出的符号的联合概率 $P(a_ib_j)$.
2. 信道传输概率 $P(b_j|a_i)$,即发送为 a_i,通过信道传输接收到 b_j 的概率. 这描述了信道噪声特性. 又称为前向概率.
3. 先验概率 $P(a_i)$. 指接收到一个输入 a_i 以前的概率.
4. 后验概率 $P(a_i|b_j)$. 指接收到一个输出符号 b_j 后的输入符号 a_i 的概率.

上述概率之间有如下关系:

1. $P(a_ib_j)=P(a_i)P(b_j|a_i)=P(b_j)P(b_j|a_i)$.
2. $P(b_j)=\sum_{i=1}^{r}P(a_i)P(b_j|a_i)\quad (j=1,2,\cdots,s)$.
3. $P(a_i|b_j)=P(a_i)\dfrac{P(b_j|a_i)}{P(b_j)}\quad (P(b_j)\neq 0)$且

$$\sum_{i=1}^{r}P(a_i|b_j)=1\quad (j=1,2,\cdots,s).$$

离散无记忆扩展信道(discrete memoryless expanding channel)　离散信道的一种类型. 一般离散无记忆信道的输入和输出均是一系列的离散随机变量,即为随机序列. 输入或输出的随机序列中每一随机变量都取值于同一输入或输出的符号集. 对于离散无记忆信道的 N 次扩展信道,若信源也是无记忆的,则有

$$I(X;Y)=\sum_{i=1}^{N}I(X_i;Y_i).$$

一般情况下,信道的输入序列 $X=(X_1,X_2,\cdots,X_N)$中的随机变量 $X_i(i=1,2,\cdots,N)$取自同一信源符号集 A,并具有同一种概率分布,而且通过相同的信道传递到输出端,即有

$$I(X_1;Y_1)=I(X_2;Y_2)=\cdots=I(X_N;Y_N)=I_N(X;Y),$$

式中 N 为序列的长度. 则离散无记忆信道的 N 次扩展信道的平均互信息量等于原来信道的平均互信息量 $I_N(X;Y)$的 N 倍,即 $I(X;Y)=NI_N(X;Y)$.

离散无噪无损信道(discrete noiseless lossness channel)　离散信道的一种类型. 若离散信道的输入和输出符号之间有确定的一一对应关系,即 $y=f(x)$,其信道传输概率为

$$P(y|x)=\begin{cases}1 & (y=f(x)),\\ 0 & (y\neq f(x)),\end{cases}$$

即有

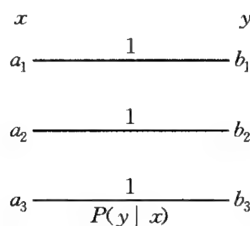
$$P(b_j|a_i)=P(a_i|b_j)=\begin{cases}0 & (i\neq j),\\ 1 & (i=j)\end{cases}$$

$$(i,j=1,2,3,\cdots),$$

这时信道的疑义度(损失熵 $H(X|Y)$)和信道的噪声熵 $H(Y|X)$都等于零. 故这类信道的平均互信息量为

$$I(X;Y)=H(X)=H(Y).$$

这种输入和输出有确定的一一对应关系的信道称为无噪无损信道. 该信道模型如图所示.



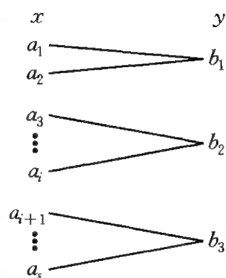
离散有噪无损信道(discrete noisy lossness channel)　离散信道的一种类型. 若信道输入一个 x 值对应有几个输出 y 值,而且每个 x 值对应的 y 值不重合. 即在这类信道中,输入符号通过信道传输可能有若干个输出符号. 输入与输出不是一一对应关系,而输出符号可分成互不相交的一些集合. 这种信道的损失熵 $H(X|Y)=0$,而噪声熵$H(Y|X)\neq 0$,则信道的平均信息量为

$$I(X;Y)=H(X)<H(Y).$$

称这类信道为有噪无损信道. 该信道模型如图所示.

离散无噪有损信道(discrete noiseless loss

channel） 离散信道的一种类型.若信道的输出 y 是输入 x 的确定函数，但不是一一对应关系，而是多一对应关系，前向概率 $P(y|x)$ 等于 0 或 1，后验概率 $P(x|y)$ 不等于 0 或 1，这类信道的噪声熵 $H(Y|X)=0$，而损失熵 $H(X|Y)\neq 0$. 即信道接收到符号 y 后不能完全消除对 x 的不确定性，信息有损失，但输出 y 的平均不确定性没有增加，它的平均互信息量为



$$I(X;Y)=H(Y)<H(X).$$

称这类信道为无噪有损信道，该信道模型如图所示.

对称离散信道(symmetric discrete channel) 离散信道的一种类型.离散信道中有一类特殊的信道，其特点是信道矩阵具有很强的对称性.所谓对称性，就是指信道矩阵 $[P]$ 中每一行都是由同一 $\{P'_1, P'_2, \cdots, P'_s\}$ 集的不同诸元素排列组成，并且每一列也都是由 $\{q'_1, q'_2, \cdots, q'_r\}$ 集的不同诸元素排列组成，具有这种特性的信道称为对称离散信道.例如，

$$[P_1]=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3}\end{bmatrix},$$

$$[P_2]=\begin{bmatrix}\frac{1}{2} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2}\end{bmatrix},$$

可见 $[P_1]$ 和 $[P_2]$ 均为对称离散信道.对称离散信道的信道容量为

$$\begin{aligned}C&=\max_{P(x)}[H(Y)-H(P'_1,P'_2,\cdots,P'_s)]\\ &=\lg s-H(P'_1,P'_2,\cdots,P'_s)\ (\text{比特/符号}).\end{aligned}$$

由此式可见信道容量只与对称信道矩阵中任一行矢量 $\{P'_1, P'_2, \cdots, P'_s\}$ 有关.

均匀离散信道(uniform discrete channel) 离散信道的一种类型.对于对称离散信道，若输入符号和输出符号的个数相同(如都等于 r)，信道矩阵为

$$[P]=\begin{bmatrix}\overline{P} & \frac{P}{r-1} & \cdots & \frac{P}{r-1}\\ \frac{P}{r-1} & \overline{P} & \cdots & \frac{P}{r-1}\\ \vdots & \vdots & & \vdots\\ \frac{P}{r-1} & \frac{P}{r-1} & \cdots & \overline{P}\end{bmatrix},$$

其中 $P+\overline{P}=1$，称具有上述特性的信道为均匀信道或强对称信道.这类信道中总的错误传输概率为 P，对称地平均分配给 $r-1$ 个输出符号.正确传输概率为 $\overline{P}$，信道矩阵中各列之和等于 1，它是信道的一类特例.均匀信道的信道容量为

$$C=\lg r-P\lg(r-1)-H(P,1-P)\ (\text{比特/符号}).$$

信道矩阵(channel matrix) 一般离散单符号信道的传输概率的矩阵形式，即

$$\begin{array}{c}\\ a_1\\ a_2\\ \vdots\\ a_r\end{array}\begin{array}{c} \begin{array}{cccc}b_1 & b_2 & \cdots & b_s\end{array}\\ \begin{bmatrix}P(b_1|a_1) & P(b_2|a_1) & \cdots & P(b_s|a_1)\\ P(b_1|a_2) & P(b_2|a_2) & \cdots & P(b_s|a_2)\\ \vdots & \vdots & & \vdots\\ P(b_1|a_r) & P(b_2|a_r) & \cdots & P(b_s|a_r)\end{bmatrix}\end{array}.$$

为表达方便，可写成 $P(b_j|a_i)=P_{ij}$. 于是有

$$[P]=\begin{bmatrix}P_{11} & P_{12} & \cdots & P_{1s}\\ P_{21} & P_{22} & \cdots & P_{2s}\\ \vdots & \vdots & & \vdots\\ P_{r1} & P_{r2} & \cdots & P_{rs}\end{bmatrix},$$

而且满足 $P_{ij}\geqslant 0$ (i 是行标号，j 是列标号)及

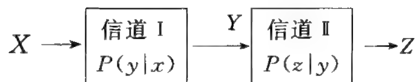
$$\sum_{j=1}^{s}P_{ij}=1\quad(i=1,2,\cdots,r).$$

该矩阵完全描述了信道的特性，称其为信道矩阵.

信道疑义度(channel equivocation) 信息传输理论的基本概念之一.在输出端收到输出符号 Y 后，对于输入端的符号 X 尚存在不确定性，即存在疑义.对 X 尚存在的不确定性是由于干扰(噪声)引起的，信道的这种特性称为信道疑义度.它可用条件熵 $H(X|Y)$ 来表征，即

$$\begin{aligned}H(X|Y)&=\sum_{x\cdot y}P(xy)\frac{1}{P(x|y)}\\ &=\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_ib_j)\frac{1}{P(a_i|b_j)}.\end{aligned}$$

串联信道的互信息量(mutual information of series channel) 信息传输理论的基本概念之一.设有一离散单符号信道 Ⅰ，其输入符号集为 $A=\{a_1, a_2, \cdots, a_r\}$，输出符号集为 $B=\{b_1, b_2, \cdots, b_s\}$，并设有另一离散单符号信道 Ⅱ，其输入符号集为 B，输出符号集为 $C=\{c_1, c_2, \cdots, c_t\}$. 将这两个信道串联起来，如下图所示.



$$\sum_{y}P(y|x)=1,\quad \sum_{z}P(z|y)=1.$$

图中 X, Y, Z 均为随机变量，$P(y|x)$ 为信道 Ⅰ 的传递概率，$P(z|y)$ 是信道 Ⅱ 的传递概率.则有

$$I(XY;Z)\geqslant I(Y;Z),$$

当且仅当 $P(z|xy)=P(z|y)$ 时等式成立.上式表明，当接收到 Z 后获得关于联合变量 X 和 Y 的信息量 $I(XY;Z)$ 大于等于接收到 Z 后获得的关于变量

Y 的信息量 $I(Y;Z)$. 称 $I(XY;Z)$ 为串联信道的互信息量.

数据处理定理(data processing theorem) 信息传输理论的基本定理之一. 在串联信道中, $P(z|xy)=P(z|y)$, 这表示随机变量 Z 只依赖于变量 Y 而与前面的变量 X 无直接关系, 故随机变量 X,Y,Z 组成一个马尔可夫链, 因此有

$$I(X;Z) \leqslant I(X;Y), \quad I(X;Z) \leqslant I(Y;Z).$$

上式称为数据处理定理, 它表示通过串联信道的传输只会丢失更多的信息. 在任何信息传输系统中, 最后获得的信息至多是信源所提供的信息. 这就是信息不增性原理.

信息传输速率(transmission rate of information) 信息传输理论的基本概念之一. 信道中平均每个符号所能传送的信息量, 用 R 表示. 因平均互信息量 $I(X;Y)$ 即为接收到符号 Y 后平均每个符号获得的关于 X 的信息量, 故信道的信息传输速率就是平均互信息量, 即

$$R = I(X;Y) = H(X) - H(X|Y).$$

定义单位时间内平均传输的信息量为

$$R_t = \frac{1}{t}I(X;Y) = \frac{1}{t}H(X) - \frac{1}{t}H(X|Y),$$

称 R_t 为信息传输速率, 信息传输速率的单位为比特/秒或奈特/秒.

信道容量(capacity of channel) 信道无错误传送理论上的最大信息率. 设信道的输入集为 $x\in A$, 输出集为 $y\in B$. 信道的条件概率为 $P(y|x)$. 当 A 和 B 都是离散集时, 信道所传送的信息率就是输出符号和输入符号之间的交互信息, 即

$$I(x,y) = \sum_{x\in A}\sum_{x\in B}P(x)P(y|x)\lg\frac{P(y|x)}{\sum_{x\in A}P(x)P(y|x)}.$$

信道的交互信息不但依赖于信道的特性, 还依赖于输入符号的概率 $P(x)$. 若能由改变编码器来调整 $P(x)$, 使交互信息 $I(x,y)$ 最大, 就能充分利用信道传输信息的能力, 这个最大值即为该信道的信道容量 C,

$$C = \max_{P(x)} I(x,y).$$

计算信道容量实际上就是一个求交互信息最大值的问题. 当 A 或 B 是连续集时, 相应的概率 $P(x)$ 应理解为概率密度.

离散信道平均互信息量极值定理(extremal theorem of the average mutual information of discrete channel) 信息传输理论的基本定理之一. 该定理断言: 离散信道的平均互信息量 $I(X;Y)$ 达到极大值(即等于信道容量)的充分必要条件是输入概率分布 $\{P_i\}$ 满足:

1. 对所有 x_i, $I(x_i;Y)=C$, 其 $P_i\neq 0$.
2. 对所有 x_i, $I(x_i;Y)\leqslant C$, 其 $P_i=0$.

式中 C 为信道容量.

离散信道平均互信息量极值定理说明: 当信道平均互信息量达到信道容量时, 输入信源符号集中每一个信源符号 x_i 对输出端 Y 提供相同的互信息量, 只是概率为零的除外.

离散无记忆扩展信道的信道容量(channel capacity of discrete memoryless expanding channel) 一种特殊的信道容量. 对于一般的离散无记忆信道的 N 次扩展信道, 其平均互信息量为

$$I(X;Y) = \sum_{i=1}^{N} I(X_i;Y_i),$$

其信道容量为

$$C^N = \max_{P(x)} I(X;Y) = \max_{P(x)}\sum_{i=1}^{N} I(X_i;Y_i) = \sum_{i=1}^{N} C_i,$$

式中

$$C_i = \max_{P(x)} I(X_i;Y_i)$$

是某时刻 i 通过离散无记忆信道传输的最大信息量. 由于输入随机序列 $X=(X_1,X_2,\cdots,X_i,\cdots,X_N)$ 在同一信道中传输, 故 $C_i=C(i=1,2,\cdots,N)$, 即 $C^N=NC$. 此式说明离散无记忆 N 次扩展信道的信道容量等于原单位符号离散信道容量的 N 倍, 且只有输入信源是无记忆的和每一输入变量 X_i 的分布各自达到最佳分布 $P(x)$ 时, 才能达到这个信道容量 NC.

信源与信道的匹配(matching of source and channel) 信息传输理论的基本概念之一. 信源发出的消息(符号)需要通过信道来传输. 对于某一给定的信道, 其信道容量是一定的, 而且只有当输入符号的概率分布 $P(x)$ 满足一定条件时才能达到信道容量 C. 即只有一种信源才能使某一信道的信息传输速率达到最大. 一般信源与信道连结时, 其信息传输速率 $R=I(X;Y)$, 并未达到最大. 这样, 信道的信息传输速率还有提高的可能. 若某一信源通过信道传输时, 信道的信息传输速率达到了信道容量, 则称此信源与信道达到匹配状态.

信道剩余度(redundancy of channel) 信息传输理论的基本概念之一. 若某一信源通过信道传输时, 信源与信道未达到匹配状态, 则认为信道有剩余, 称为信道剩余度. 定义为

$$信道剩余度 = C - I(X;Y),$$

式中 C 是该信道的信道容量, $I(X;Y)$ 是信源通过该信道实际传输的平均信息量. 由此关系还可定义信道的相对剩余度.

$$信道相对剩余度 = \frac{C - I(X;Y)}{C}$$

$$=1-\frac{I(X;Y)}{C}.$$

在无损信道中，$C=\lg r$（r 为信道输入符号的个数），而 $I(X;Y)=H(X)$（输入信道的信源熵），因而

$$\text{无损信道的相对剩余度}=1-\frac{H(X)}{\lg r}.$$

该式即为信源的剩余度．对于无损信道，可以通过信源编码来减少信源的剩余度，使信源与信道达到匹配状态．

信源编码(source coding)　信息传输理论的基本概念之一．编码的实质是对信源的原始符号按一定的数学规则进行的一种变换．其目的是在不失真或允许一定失真条件下，如何用尽可能少的符号来传送信源信息，以便提高信息传输率．对于无失真信源编码的数学描述如下图．它的输入是信源符号集

$$S=\{s_1,s_2,\cdots,s_q\}\longrightarrow\boxed{\text{编码器}}\longrightarrow C=\{w_1,w_2,\cdots,w_q\}$$
$$\uparrow$$
$$A=\{a_1,a_2,\cdots,a_r\}$$

$S=\{s_1,s_2,\cdots,s_q\}$．同时存在着另一符号集 $A=\{a_1,a_2,\cdots,a_r\}$，集中的元素 a_j 是适合信道传输的，称为码元．编码器的任务是将信源符号集中的符号 s_i（或长为 N 的信源符号序列）变换成由 $a_j(j=1,2,\cdots,r)$ 组成的长度为 l_i 的一一对应的序列，这种码元序列称为码字 w_i．所有这些码字的集合称为码 C，即

$$C=\{w_1,w_2,\cdots,w_q\}.$$

等长码(fixed-length code)　信源编码的一种类型．若码中所有码字含有的码元个数（即码长）都相同，则称此码为等长码，反之称变长码．若码中含有的所有码字都不相同，则称此码为单值码，反之称多值码．假定有一码 C，它把信源符号集 S 中的符号 s_i 一一变换成码 C 中的码字 w_i，则码 C 的 N 次扩展码是指 N 次扩展信源 S^N 中信源符号序列所对应的原编码码字的序列构成的码．若等长码是单值码，则它的任意有限长 N 次扩展码也是单值码．对信源 S 进行等长编码时，必须满足

$$q\leqslant r^l,$$

式中 l 为等长码的码长，q 为信源符号集 S 中信源符号的个数，r 为符号集 A 中码元的个数．

惟一可译码(unique decodable code)　信源编码的一种类型．若要实现无失真编码，这不但要求信源符号 $s_i(i=1,2,\cdots,q)$ 与码字 $w_i(i=1,2,\cdots,q)$ 是一一对应的，而且要求码元序列的反变换也是惟一的，即任意一串有限长的码元序列只能被惟一地译成所对应的信源符号序列，这种码称为惟一可译码．若要所编的码是惟一可译码，则要求任意有限长的信源符号序列所对应的码元序列各不相同，即要求码的任意有限长 N 次扩展码都是单值码．若对信源 S 的 N 次扩展信源进行等长编码时，必须满足

$$q^N\leqslant r^l,$$

才能使所编得的等长码为惟一可译码．对上式两边取对数，则有

$$N\lg q\leqslant l\lg r\ \text{或}\ \frac{l}{N}\geqslant\frac{\lg q}{\lg r}.$$

变长码(variable-length code)　信源编码的一种类型．若码中码字含有的码元个数（即码长）不同，则称此码为变长码．应用变长码往往在 N 不很大时即可编出效率很高而且无失真的码．变长码必须是惟一可译码，才能实现无失真编码．为此它本身必须是单值码，且其任意有限长 N 次扩展码也都必须是单值码．

惟一可译变长码存在的条件：设信源符号集为 $S=\{s_1,s_2,\cdots,s_q\}$，码元集 $A=\{a_1,a_2,\cdots,a_r\}$，又设码字为 $w_1,w_2,\cdots,w_q$，其分别对应的码长为 $l_1,l_2,\cdots,l_q$，则存在惟一可译码的充分必要条件是

$$\sum_{i=1}^{q}r^{-l_i}\leqslant 1.$$

最佳码(optimal code)　信源编码的一种类型．对于某一信源和某一码元集，若有一个惟一可译码，其平均长度 $\bar{L}$ 小于等于所有其他惟一可译码的平均长度，则称该码为最佳码或紧致码．无失真信源编码的基本问题就是寻找最佳码．若一个离散无记忆信源 S 具有熵为 $H(S)$、并有码元集 $A=\{a_1,a_2,\cdots,a_r\}$，则总可找到一种无失真编码方法，构成惟一可译码，使其平均码长 $\bar{L}$ 满足

$$\frac{H(S)}{\lg r}\leqslant\bar{L}\leqslant 1+\frac{H(S)}{\lg r}.$$

上式表示最佳码的平均长度的下限值与信源熵 $H(S)$ 成正比．

r 进制霍夫曼码(r-symbol Huffman code)　信源编码的一种类型．二进制霍夫曼码的编码方法可以推广到任意 r 进制编码中．不同的仅是每次把 r 个元（概率最小的）合并成一个新的信源符号，并分别用 $0,1,\cdots,(r-1)$ 等码元表示．为了使短码字得到充分利用，使平均码长为最短，必须使最后一步的缩减信源有 r 个信源符号．因此对于 r 进制霍夫曼编码，信源 S 的符号个数 q 必须满足

$$q=(r-1)Q+r,\tag{1}$$

其中 Q 表示缩减的次数，$(r-1)$ 为每次缩减所减少的信源符号个数．对于 r 进制霍夫曼码，q 为任意正整数时不一定能找到一个 Q 满足上式．为此设一些信源符号并使它们对应的概率为零，即 $s_{q+1},s_{q+2},\cdots,s_{q+t}$ 对应的概率

$$P_{q+1}=P_{q+2}=\cdots=P_{q+t}=0,$$

此时 $q+t$ 便能满足(1)式．这样得到的 r 进制霍夫曼码一定是最佳码．

法诺码(Fano code) 信源编码的一种类型.法诺码的编码程序如下:首先,将信源符号以概率递减的次序排列起来,将排列好的信源符号划分成两大组,使每组的概率和近于相同,并各赋于一个二进制码元"0"和"1".然后,将每一大组的信源符号再分成两组,使同一组的两个小组的概率和近于相同,并分别赋于一个二进制码元.依次下去,直至每个小组只剩一个信源符号为止.这样,信源符号对应的码元序列则为编得的码字.法诺编码不是最佳编码方法,所以不是一种最佳码.

等长信源编码定理(source coding theorem of fixed-length) 信源编码理论的基本定理之一.一个熵为 $H(S)$ 的离散无记忆信源,若对信源长为 N 的符号序列进行等长编码,设码字是从 r 个字母的码元集中选取 l 个字母组成.对于任意 $\varepsilon>0$,只要满足

$$\frac{l}{N} \geqslant \frac{H(S')+\varepsilon}{\lg r}, \tag{1}$$

则当 N 足够大时,几乎可实现无失真编码,即译码错误能为任意小.反之若

$$\frac{l}{N} \leqslant \frac{H(S')-\varepsilon}{\lg r},$$

则不可能实现无失真编码,而当 N 足够大时,译码错误概率近似等于1.

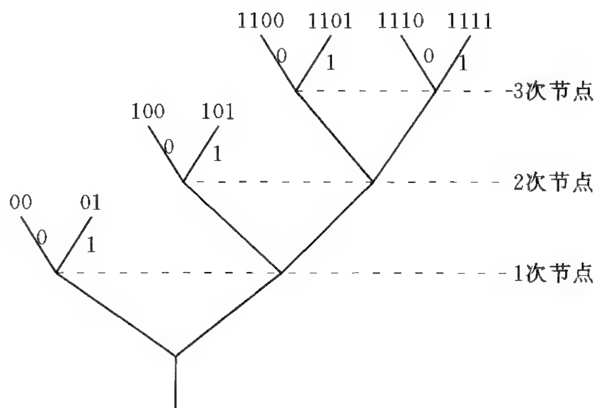
(1)式可改写为 $l\lg r>NH(S')$,该式左边表示长为 l 的码字能载荷的最大信息量,而右边表示长为 N 的信源符号序列平均携带的信息量.故该定理表明:只要码字传输的信息量大于信源符号序列携带的信息量,总可实现无失真编码,且传输效率接近1,其必要条件为 N 充分大.

克拉夫特不等式(Kraft inequality) 信源编码理论中的一个重要不等式.当一个码的任意码字与比它更长的任意码字的字首不相同时,称此码为满足字首条件的码.由码字分别为 $N_i(i=1,2,\cdots,M)$ 的 M 个码字所组成而且又满足字首条件的码,其存在的充分必要条件是满足公式

$$\sum_{i=1}^{M} n^{-N_i} \leqslant 1.$$

此式称为克拉夫特不等式.

树形编码法(tree coding method) 一种编码方法.将码字的全体集合 $C=\{x_1,x_2,\cdots,x_m\}$ 简称码.将码的全体形象地用树表示出来,称为树形编码法.如下图所示,各码字分配在树的终端节点上.由树根出发到一次节点的两个枝叉,分别对应于码字的第1个码元(0或1).一般由 k 次节点分出的两个枝叉对应于码字的第 $k+1$ 个码元,当字母表中的文字数目为 $n\geqslant 2$ 时,则由节点伸出的不同枝叉的数目也应为 n 个.



变长信源编码定理(source coding theorem of variable length) 信源编码理论的基本定理之一.对于由 n 个字母构成的字母表 $A=\{a_1,a_2,\cdots,a_n\}$,编码由 m 个消息构成的信源 X,

$$X=\begin{Bmatrix} u_1 & u_2 & \cdots & u_m \\ P_1 & P_2 & \cdots & P_m \end{Bmatrix},$$

式中 $u_1,u_2,\cdots,u_m$ 为消息,$P_1,P_2,\cdots,P_m$ 为相应的消息发生的概率,对应于各消息的码字的平均长度 $\overline{N}$ 满足不等式

$$\frac{H(X)}{\lg n} \leqslant \overline{N} < \frac{H(X)}{\lg n}+1.$$

即对任意的单义可译码,其平均码字长一定满足下列不等式:

$$\overline{N} \geqslant \frac{H(X)}{\lg n}.$$

此定理称为无噪声信道变长信源编码定理.

仙农第一定理(Shannon first theorem) 信源编码理论的基本定理之一.即变长无失真信源的编码定理.离散无记忆信源 S 的 N 次扩展信源 $S^N=\{u_1,u_2,\cdots,u_{qN}\}$,其熵为 $H(S^N)$,并有码元集 $A=\{a_1,a_2,\cdots,a_r\}$.对信源 S^N 进行编码,总可以找到一种编码方法,构成惟一可译码,使信源 S 中每个信源符号所需的码字平均长度满足:

$$\frac{H(S)}{\lg r}+\frac{1}{N} > \frac{\overline{L}_N}{N} \geqslant \frac{H(S)}{\lg r}$$

或

$$H_r(S)+\frac{1}{N} > \frac{\overline{L}_N}{N} \geqslant H_r(S).$$

当 $N\to\infty$ 时,则

$$\lim_{N\to\infty}\frac{\overline{L}_N}{N}=H_r(S),$$

式中

$$\overline{L}_N=\sum_{i=1}^{q^N} P(u_i)\lambda_i,$$

而 λ_i 是 u_i 所对应的码字长度.$\overline{L}_N/N$ 的含义是:为了得到这个平均值,不是对单个信源符号进行编码,而是对 N 个信源符号序列 u_i 进行编码.

编码效率(coding efficiency)　信源编码理论的基本概念之一.设对信源 S 进行编码所得到的平均码长为 $\bar{L}$,因为 $\bar{L}$ 一定大于或等于 $H_r(S)$,则编码效率定义为

$$\eta=\frac{H_r(S)}{\bar{L}}\quad(\eta\leqslant 1).$$

对于同一信源,若码的平均码长 $\bar{L}$ 越短,越接近于极值 $H_r(S)$,则信息传输速率就越高,就越接近于无噪无损信道容量,这时 η 也就越接近于 1.可见用编码效率来衡量各种各样编码的优劣是合适的.为了衡量各种编码与最佳码的差距,定义码的剩余度为

$$1-\eta=1-\frac{H_r(S)}{\bar{L}}.$$

译码规则(decoding rule)　译码过程中必须遵循的规定.设信道的输入码字集为 $A=\{a_i\}(i=1,2,\cdots,r)$,输出信号集为 $B=\{b_j\}(j=1,2,\cdots,s)$.制定译码规则就是设计一个函数 $F(b_j)$,它对于每一个输出信号 b_j 确定一个惟一的输入码字 a_i 与其对应(单值函数),即

$$F(b_j)=a_i\quad(i=1,2,\cdots,r;j=1,2,\cdots,s).$$

当确定 $F(b_j)=a_i$ 后,若信道输出端接收到的信号为 b_j,而发送的码字不是 a_i,则认为有错.其错误概率 $P(e|b_j)$ 称为条件错误概率.而译码的条件正确概率为

$$P[F(b_j)|b_j]=P(a_i|b_j).$$

两者的关系为

$$P(e|b_j)=1-P(a_i|b_j)=1-P[F(b_j)|b_j].$$

经过译码后的错误概率 P_E 应是条件错误概率对 B 集取平均值,即

$$P_E=\sum_{j=1}^{s}P(b_j)P(e|b_j).$$

称 P_E 为平均错误概率.

最小错误概率准则(criterion of minimum error-probability)　最小译码错误概率的标准.若选择译码函数为 $F(b_j)=a^*$,并使之满足条件

$$P(a^*|b_j)\geqslant P(a_i|b_j)(\text{对所有 } i),$$

则使条件错误概率 $P(e|b_j)$ 为最小.由译码函数 $F(b_j)=a^*$ 所确定的这个不等式称为最小错误概率准则.

法诺不等式(Fano inequality)　编码理论中的一个重要不等式.平均错误概率 P_E 与译码函数有关,而译码函数又由信道特性来决定.由于信道中存在噪声,导致输出端发生错误,可见 P_E 与信道疑义度 $H(X|Y)$ 有关.

$$H(X|Y)\leqslant H(P_E,1-P_E)+P_E\lg(r-1),$$

称该式为法诺不等式.式中 $H(P_E,1-P_E)$ 为平均错误概率 P_E 的熵.法诺不等式说明接收到 Y 后关于 X 的平均不确定性 $H(X|Y)$ 可分为两部分:第一部分是指接收到 Y 后是否产生错误的不确定性 $H(P_E,1-P_E)$;第二部分表示当错误发生后,到底是发送哪个输入符号而造成错误的不确定性

$$P_E\lg(r-1).$$

汉明距离(Hamming distance)　编码理论的基本概念之一.长度为 n 的两个二进制序列(码字)α_i 和 β_j 之间的距离是指 α_i 和 β_j 之间对应位置上码元取值不同的个数,用符号 $D(\alpha_i,\beta_j)$ 表示.这种码间距离通常称为汉明距离.例如,$\alpha_i=101111$,$\beta_j=111100$,则 $D(\alpha_i,\beta_j)=3$.一般若令

$$\alpha_i=(\alpha_{i_1},\alpha_{i_2},\cdots,\alpha_{i_n}),\quad\beta_j=(\beta_{j_1},\beta_{j_2},\cdots,\beta_{j_n}),$$

则 α_i 与 β_j 的汉明距离为

$$D(\alpha_i,\beta_j)=\sum_{k=1}^{n}\alpha_{i_k}\oplus\beta_{j_k},$$

式中 $\oplus$ 为逻辑导式运算符号模 2 加.在二进制码 C 中,任意两个码字的汉明距离的最小值称为该码 C 的最小距离,即

$$D_{\min}=\min\{D(C_i,C_j)\}\quad(C_i\neq C_j,C_i,C_j\in C).$$

在任一码中,码的最小距离与该码的译码平均错误概率 P_E 有关,$D_{\min}$ 越大,P_E 越小.可见在选择编码规则时,码字之间的最小距离越大越好.

仙农第二定理(Shannon second theorem)　有噪信道编码的定理.设某信道有 r 个输入符号、s 个输出符号,信道容量为 C.当信息传输速率 $R<C$ 时,只要码长 n 足够长,总可以在输入的集合中找到 M 个码字(代表 M 个等概率分布的消息)组成的一个码 $M=2^{nR}\leqslant 2^{n(C-\varepsilon)}$($\varepsilon$ 是任意小的正数)和它相应的译码规则,使信道输出的平均错误概率 P_E 任意小.定理说明在满足条件 $R<C$ 时平均错误概率趋于零的编码是存在的.

仙农第二定理的逆定理(converse theorem of Shannon second theorem)　有噪信道编码定理的逆定理.设某信道有 r 个输入符号、s 个输出符号,信道容量为 C.令 ε 为任意小的正数.若选用码字个数 $M\geqslant 2^{n(C+\varepsilon)}$,则无论码长 n 多大也不能编出使译码后信道输出的平均错误概率 P_E 任意小的码.定理表明:当选择 $M\geqslant 2^{n(C+\varepsilon)}$ 时,即信息传输速率 $R\geqslant C+\varepsilon$,表示信息传输速率大于信道容量,因此,在这种情况下实现无错误传输消息是不可能的.

仙农第三定理(Shannon third theorem)　保真度准则下的信源编码定理.设 $R(D)$ 为一离散无记忆信源的信息率失真函数,并且具有有限的失真度.对于任意的 $D\geqslant 0$,$\varepsilon>0$,以及任意足够长的码长 k,则一定存在一种信源编码 C,其码字个数为

$$M\leqslant e^{\{k[R(D)+\varepsilon]\}}.$$

而编码后的平均失真度为

$$d(C) \leqslant D + \varepsilon.$$

定理表示：对于任何失真度 $D \geqslant 0$，只要码长 k 足够长，总可以找到一种编码 C，使编码后平均每个信源符号的信息传输速率为

$$R' = \frac{\lg M}{k} \leqslant R(D) + \varepsilon.$$

而码的失真度为 $d(C) \leqslant D$.

仙农第三定理的逆定理(converse theorem of Shannon third theorem)　保真度准则下的信源编码定理的逆定理. 不存在平均失真度为 D，而平均信息传输速率 $R' < R(D)$ 的任何信源码. 换言之，对任意码长为 k 的信源码 C，若码字个数 $M < e^{kR(D)}$，一定有 $d(C) > D$. 该定理断言：如果编码后平均每个信源符号的信息传输速率 R' 小于信息率失真函数 $R(D)$，就不能在保真度准则下再现信源的消息.

连续信源(continuous source)　信源的一种类型. 有的信源输出的消息的取值是连续的，即可能出现的消息数是不可数的无限值，可用连续型随机变量来描述这些消息. 这种信源称为连续信源，其数学模型为在 (a,b) 或 R 中取值，概率密度函数 $p(x) \geqslant 0$，并满足

$$\int_a^b p(x)\mathrm{d}x = 1 \quad 或 \quad \int_R p(x)\mathrm{d}x = 1$$

的连续型随机变量 X 或随机过程 $X(t)$. 其中 R 表示实数集 $(-\infty, +\infty)$. 连续信源有如下特征：

1. 连续信源中消息数是无限的. 它输出的每个可能消息是随机过程 $X(t)$ 中的一个样本函数，而随机过程 $X(t)$ 中有无限多个样本函数.

2. 对于连续信源，可用有限维概率密度函数族以及与各维概率密度函数有关的统计量来描述，即 $p_1(x_1,t_1), p_2(x_1,x_2,t_1,t_2), \cdots, p_n(x_1,x_2,\cdots,x_n,t_1,t_2,\cdots,t_n)$，其中 $x_1,x_2,\cdots,x_n$ 是随机过程 $X(t)$ 在 $t_1,t_2,\cdots,t_n$ 时刻的取值，而时间 t 是连续的.

连续信源熵(entropy of continuous source)　连续信源的一种量度. 连续信源基本数学模型为

$$X: \begin{bmatrix} R \\ p(x) \end{bmatrix},$$

并满足

$$\int_R p(x)\mathrm{d}x = 1,$$

其中 R 表示全体实数，是连续变量 X 的取值范围. 若将连续变量 X 用离散变量来近似，则连续信源 X 经量化后成为离散信源，其离散信源的熵为

$$H_n(X) = -\sum_{i=1}^{n} p(x_i)\Delta \lg p(x_i) - \sum_{i=1}^{n} p(x_i)\Delta \lg \Delta,$$

式中 n 为量化数，Δ 为量化区间宽度，x_i 为第 i 个量化区间中的一个值. 当 $n \to \infty, \Delta \to 0$ 时，若极限存在，则连续信源的绝对熵为

$$\begin{aligned}\lim_{n\to\infty} H_n(X) &= -\lim_{\Delta\to 0}\sum_{i=1}^{n} p(x_i)\Delta \lg p(x_i) \\ &\quad - \lim_{\Delta\to 0}\sum_{i=1}^{n} p(x_i)\Delta \lg \Delta \\ &= -\int_a^b p(x)\lg p(x)\mathrm{d}x - \lim_{\Delta\to 0}\lg\Delta.\end{aligned}$$

定义

$$h(X) = -\int_R p(x)\lg p(x)\mathrm{d}x$$

为连续信源的微分熵，也称连续信源熵.

连续信源最大熵定理(maximum entropy theorem of continuous source)　信息理论中的一个基本定理. 若信源 $X = (X_1, X_2, \cdots, X_n)$ 是 n 维随机向量，其概率密度函数 $p(x_1,x_2,\cdots,x_n)$ 满足：

$$p(x_1,x_2,\cdots,x_n) \geqslant 0,$$

$$\int_{R^n} p(x_1,x_2,\cdots,x_n)\alpha x_1 \alpha x_2 \cdots \alpha x_n \mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n = 1,$$

$$\int_{R^n} p(x_1,x_2,\cdots,x_n) x_i x_j \mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n = K_{ij}$$

$$(1 \leqslant i,j \leqslant n),$$

则最大熵概率密度函数为正态概率密度函数，即

$$p(x_1,x_2,\cdots,x_n) = \frac{1}{\sqrt{(2\pi)^n |K|}}\exp\left[-\frac{1}{2}xK^{-1}x'\right],$$

其中 $x = (x_1,x_2,\cdots,x_n)$，x' 为 x 的转置，$K = [K_{ij}]$ 为 X 的协方差矩阵. 信源 X 的最大熵为

$$\max_p H(X) = \frac{1}{2}\lg(2\pi e)^n |K|,$$

即最大熵定理. 上式中的 $|K|$ 为协方差矩阵 K 的行列式. 最大是对所有满足上述条件的概率密度函数 $p(x_1,x_2,\cdots,x_n)$ 而取.

判别信息量(distinguishing information)　信息科学的一种信息度量. Kullback 推广了熵的概念，提出用于母体判别的差别信息量. 设 $p_1(x)$，$p_2(x)$ 为两个概率密度函数，则 $p_1(x)$ 对 $p_2(x)$ 的判别信息量定义为

$$K(p_1,p_2) = \int_{-\infty}^{\infty} p_1(x)\lg\frac{p_1(x)}{p_2(x)}\mathrm{d}x.$$

设 $S_x(w)$，$S_y(w)$ 为两个高斯平稳时间序列的功率谱时，$S_x(w)$ 对 $S_y(w)$ 的判别信息量为

$$K(S_x,S_y) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{S_x(w)}{S_y(w)} - \lg\frac{S_x(w)}{S_y(w)} - 1\right]\mathrm{d}w.$$

由于 $K(p_1,p_2)$ 和 $K(S_x,S_y)$ 是非负的，因此，判别信息量可作为 p_1 和 p_2 或 S_x 和 S_y 之间的一种"距离"，应用于平稳过程的识别、分类和谱估计中. 此外，当 $p_1(x)$ 和 $p_2(x)$ 为两个离散概率分布时，也可定义

$$K(p_1,p_2)=\sum_x p_1(x)\lg\frac{p_1(x)}{p_2(x)},$$

称其为互熵或 Kullback 熵或信息散度.在信息理论中有许多应用.

连续信源的互信息量(mutual information of continuous source)　连续信源的一种信息量.两个连续信源 X 和 Y,它们的联合熵为

$$h(X,Y)=-\iint_R p(x,y)\lg p(x,y)\mathrm{d}x\mathrm{d}y.$$

其条件熵为

$$h(Y|X)=-\iint_R p(x)p(y|x)\lg p(y|x)\mathrm{d}x\mathrm{d}y,$$

$$h(X|Y)=-\iint_R p(y)p(x|y)\lg p(x|y)\mathrm{d}x\mathrm{d}y.$$

连续信源 X 和 Y 之间的平均互信息量为

$$\begin{aligned}I(X;Y)=I(Y;X)&=h(X)-h(X|Y)\\&=h(X)+h(Y)-h(X,Y)\\&=h(Y)-h(Y|X),\end{aligned}$$

且 $I(X;Y)\geqslant 0$.平均互信息量 $I(X;Y)$ 有如下性质:

1. $I(X;Y)=I(Y;X)\geqslant 0$.

2. 若 $z=f(y)$ 是任意的映射,则

$$I(X;Y)\geqslant I(X;Z).$$

3. 若 f 具有惟一的逆变换 f^{-1}(即 $z=f(y)$ 是一一对应的映射),则 $I(X;Y)=I(X;Z)$.

高斯信源熵(entropy of Gauss source)　高斯信源的一种量度.高斯信源是指信源输出的一维随机变量 X 的概率密度分布是正态分布,即

$$p(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right],$$

式中 m 是 X 的均值,σ^2 是 X 的方差.这种连续信源称为高斯信源.高斯信源的熵为

$$h(X)=\lg\sqrt{2\pi\sigma^2}+\frac{1}{2}\lg e=\frac{1}{2}\lg 2\pi e\sigma^2,$$

式中

$$\sigma^2=\int_{-\infty}^{\infty}(x-m)^2p(x)\mathrm{d}x,$$

当 $m=0$ 时,X 的方差 σ^2 等于信源输出的平均功率.

高斯信源(Gauss source)　见"高斯信源熵".

连续信道的信息传输速率(information transmission rate of continuous channel)　信息传输理论的基本概念之一.无记忆连续信道的平均互信息量为

$$I(X;Y)=h(X)-h(X|Y),$$

且有

$$I(X;Y)\leqslant\sum_{i=1}^{N}I(X_i;Y_i).$$

对于无记忆连续信源,则有

$$I(X;Y)\geqslant\sum_{i=1}^{N}I(X_i;Y_i).$$

若信道和信源均为无记忆时,则有

$$I(X;Y)=\sum_{i=1}^{N}I(X_i;Y_i),$$

其信息传输速率为

$$R=I(X;Y)=\sum_{i=1}^{N}I(X_i;Y_i),$$

在单位时间内,连续信道的信息传输速率为

$$R_t=\lim_{t\to\infty}\frac{1}{t}I(X;Y).$$

连续信道的信道容量(channel capacity of continuous channel)　信息传输理论的基本概念之一.对于一固定的连续信道,适当选择信道输入信号的概率密度函数 $p(x)$,可使连续信道的信息传输速率 R 达到最大值.称这个最大的信息传输速率为信道容量,即

$$C=\max_{p(x)}I(X;Y)=\max_{p(x)}[H(Y)-H(Y|X)].$$

它反映了该连续信道传输信息能力的大小.对于不同的连续信道,它们存在的噪声形式不同,信道带宽以及对信号的各种限制不同,所以具有不同的信道容量.

仙农信道容量公式(Shannon channel capacity formula)　信息传输理论的一个重要公式.信道的输入和输出都是取连续的一维随机变量,而且加入信道的是加性(白)高斯噪声,则称该信道为一维高斯信道.在带宽受限条件下,它的信道容量为

$$C=\max_{p(x_i)}\sum_{i=1}^{N}I(X_i;Y_i)=\frac{N}{2}\lg\left(1+\frac{P_s}{\sigma^2}\right),$$

其中 N 是随机信号的采样数,即 $N=2FT$ (F 为带宽,T 为信号持续时间).P_s 为无噪声时的信号平均功率.σ^2 是信道噪声的平均功率.令 $\sigma^2=P_n$,故单位时间的信道容量为

$$C=F\lg\left(1+\frac{P_s}{P_n}\right),$$

称该式为仙农信道容量公式.式中 P_s/P_n 称为信道的信噪比.此公式说明信道容量与信道的带宽和信噪比有关.

连续信源的信息率失真函数(information rate distortion function of continuous source)　信息理论的一个基本函数.设连续信源 U,其概率密度为 $p(u)$.接收端变量为 V,信道传递概率密度为 $p(v|u)$.U 和 V 之间的失真函数为 $d(u,v)$,则平均失真度为

$$\overline{D}=\iint_{-\infty}^{+\infty}p(u)p(v|u)d(u,v)\mathrm{d}u\mathrm{d}v.$$

信道的平均互信息量为

$$I(U;V)=h(V)-h(V|U).$$

若确定一允许失真度为 D,凡满足 $\overline{D}\leqslant D$ 的所有试验信道的集合为

$$B_D: \{p(v|u); \overline{D} \leqslant D\},$$

则连续信源的信息率失真函数为

$$R(D) = \inf\{I(U;V)\}, \ p(v|u) \in B_D,$$

式中 inf 是指下确界，它相当于离散信源 $R(D)$ 定义式中的极小值. 虽然对连续信源可能不存在极小值，但下确界是存在的.

量子信息(quantum information) 以量子态作为载体所携带的信息. 量子信息的单元称为量子比特，它是任意两态量子系统的量子态，一般地，量子比特是两个正交态的任意叠加态. 作为经典信息单元的比特可以看做量子比特的特殊情况. 量子比特的物理载体可以是光子、电子、原子核和声子等. 量子信息的物理系统遵从量子规律，量子信息的全部过程都必须用量子理论来描述，信息的传输是指量子态(即量子比特)在量子通道中的传送，信息的处理(如计算等)指对量子态进行幺正变换，信息的提取指对量子态进行量子测量. 量子信息系统基于量子特性可以实现经典信息系统所无法做到的信息功能，如量子计算机可以有效地实现大数因子分解，从而攻破广泛使用的公开密钥 RSA 体系，量子密码则可提供安全的保密通信，量子通信网络具有比现有通信网更优异的特性.

量子信息推广和完善了经典信息. 信息从经典向量子的推广，如同数学上从实数向虚数的推广一样. 量子信息除了经典信息所具有的信源与信道的概念外，还有其独有的量子纠缠.

量子编码定理(quantum coding theorem) 量子通信系统的基本定理. 量子信源编码定理：量子信源以概率 P_i 发送密度算符为 ρ_i 的量子态，

$$\rho = \sum_i P_i \rho_i$$

表示信源的总密度算符. 量子信源编码定理表明，如果所有 ρ_i 均限制为纯态，则冯・诺伊曼熵

$$S(\rho) = -tr(\rho \lg_2 \rho)$$

确定了表征这个量子系统的信息所需的最小量子比特数. 若 ρ_i 为混合态，则相对冯・诺伊曼熵

$$S(\rho) - \sum_i P_i S(\rho_i)$$

确定了所需的最小量子比特数.

量子信道编码定理：在量子信息中，根据不同的辅助资源和通信用途，量子信道有几种不同的容量. 包括：

1. 经典容量 C. 等于通过此量子信道可靠地传送经典比特的最大速率.

2. 量子容量 Q. 指通过信道完全可靠地传送量子比特的最大速率.

3. 经典辅助量子容量 Q_2. 定义为在通信双方无限制的经典通信辅助下，通过信道完全可靠地传送量子比特的最大速率.

4. 纠缠辅助经典容量 C_E. 定义为在通信双方拥有无限制的事先分享的量子纠缠的辅助下，通过此信道可靠地传送经典比特的最大速率.

对于所有已知的量子信道，这些容量之间满足关系 $Q \leqslant Q_2 \leqslant C \leqslant C_E$，这些容量各自随着信道参数的变化而独立变化，但是计算起来比较困难.

量子纠缠(quantum entanglement) 量子系统的一种性质. 在量子力学的数学表述中，量子纠缠是指含多个子系统的复合量子系统所表现出来的不可分的性质，即描述复合系统的态的密度函数不能表为子系统密度函数的经典关联.

从物理特性上看，量子纠缠就是指含多个子系统的复合量子系统中，各子系统的可观测物理量所表现出来的非局域关联性质. 量子纠缠是量子信息学尤其是量子通信的核心内容. 许多量子信息学任务都要借助具有纠缠性质的量子态来完成.

量子通信(quantum communication) 使用量子信道的一种通信系统. 它完成经典信息或量子信息的传输，以及量子纠缠的分配. 在这里，量子信道是指可以直接传递携带量子态的粒子的通道或通信各方事先分享的纠缠态.

量子隐形传态是量子通信中最为重要的内容，可以以此为基础完成许多其他的量子通信任务. 量子隐形传态过程是这样的：通信双方事先分享有效辐射功率(EPR)纠缠粒子对，发送方对携带待传送量子态的粒子和自己分享的纠缠粒子做联合的 Bell 基测量，然后把得到的测量结果(两比特的经典信息)通过经典通道告诉接收方，接收方根据此结果对自己的纠缠粒子做相应的幺正变换，即得到待送的量子态.

量子通信网络(quantum communication network) 一种采用量子通信系统的网络. 量子通信网络由众多分离的节点组成，量子信息就存储在这些节点中. 在节点处，代表量子信息的单元——量子比特可以被局域地操纵. 这些节点将量子信息通道连结起来，形成量子通信网络. 在网络内部，信息的交换可以通过传递量子比特来实现. 在实际的物理系统中，实现量子通信网络，需要被束缚的离子或原子作为节点，而量子信息通道则由光纤或者类似的光子"线路"来实现.

量子通信网络具有安全性、多端分布计算、降低通信复杂性等优点.

量子通信复杂度(quantum communication complexity) 量子通信系统的一种性能. 相距较远的通信各方分别持有某一布尔函数输入值的一部分，其目的在于共同决定函数的值，若各方均能获知正确的输出值，所需要传递的最小信息量即为通信

复杂度. 利用量子力学的一些基本特性,例如量子态的叠加性和非局域性等,通信的复杂度可以得到一定程度的降低. 对某些特殊形式的布尔函数来说,如果通信各方相互传递以量子比特表示的量子信息,则此通信会较之传递经典信息更为有效,即通信复杂度得到降低. 另一种降低通信复杂度的方法是:各方仍然交换经典信息,但是通信之前共享处于量子纠缠态的粒子对,利用纠缠特性,使得针对一些特定的布尔函数,减少所需的通信量成为可能,从而降低量子通信复杂度.

消相干(decoherence) 量子系统的一种性质. 消相干是指量子态由于与环境的相互作用引起的量子态的量子相干性(叠加性的衰退).

量子编码(quantum coding) 量子通信中的编码方式. 量子编码是用一些特殊的量子态来表示量子比特,以达到克服消相干的目的. 量子编码有量子纠错码、量子避错码和量子防错码三种形式.

量子密码(quantum cryptography) 量子通信的一种保密技术. 将欲传送的信息采取基于某种量子特性的方法进行加密,以致惟有量子信息通道的合法用户才能从中恢复出正确的信息. 目前所说的量子密码实质上是量子密钥的分配技术. 采用量子信息通道来传送密钥,其安全性由物理定律本身来保证. 因此,这种密码技术有能力抵抗强大的破译技术和计算工具的攻击. 已提出的两者之间,即点对点之间建立密钥的方案可以划分为四类:

1. 基于两种共轭基的四态方案,其代表为 1984 年由贝内特(Bennett)和布拉萨德(Brassard)提出的 BB84 协议.

2. 基于两个非正交态的两态方案,如 B92 协议;

3. 基于有效辐射功率(EPR)佯谬的 EPR 协议,是由埃克特(Ekert)于 1991 年提出的,故也称为 E91 协议.

4. 基于正交态的密钥分配方案,其安全性由正交态的不可克隆定理保证.

所有这些方案的核心在于应用到量子力学的叠加性,以及任何信息的提取(例如窃听)都会干扰信息通道. 量子密码原则上可提供不可破译、不可窃听的保密通信体系. 为了进一步走向实用化、商业化,人们又提出将量子密钥分配网络化的想法,即网络上任意两点之间可以建立密钥. 现在已经提出了树形、环形和星形的网络,这几种网络通信方案仍然以点对点的四种量子密钥分配协议为基础.

量子不可克隆定理(quantum no-cloning theorem) 量子力学中的一个基本定理. 根据量子力学的叠加原理及线性性可以证明,不存在任何操作(包括幺正演化和测量)能够将完全未知的量子态精确克隆. 这一定理于 1982 年由伍特斯(Wootters)和泽勒克(Zurek)提出,其简要证明如下:

设存在一物理过程 U(包括幺正演化和测量)可以克隆完全未知的量子态,对于任意两个量子态 $|\Psi_0>,|\Psi_1>(|\Psi_0>\neq|\Psi_1>)|$,必有

$$U|\Psi_0>|0>=|\Psi_0>|\Psi_0>,$$
$$U|\Psi_1>|0>=|\Psi_1>|\Psi_1>.$$

由于量子力学是线性的,对于叠加态 $|\Psi>=c_0|\Psi_0>+c_1|\Psi_1>$,有

$$\begin{aligned}U|\Psi>|0>&=U(c_0|\Psi_0>|0>+c_1|\Psi_1>|0>)\\&=c_0|\Psi_0>|\Psi_0>+c_1|\Psi_1>|\Psi_1>\\&\neq|\Psi>|\Psi>.\end{aligned}$$

由此,定理得证.

普适量子克隆(universal quantum cloning) 量子力学中的一种现象. 在不违背量子不可克隆定理的前提下,对于完全未知的量子态,可以进行不完全精确的确定性克隆,克隆的精确度由初态 $|\Psi_0>$ 和末态 $|\Psi_1>$ 之间的保真度 $F=|<\Psi_0|\Psi_1>|^2$ 描述. 如果对于任意量子态 $|\Psi>$ 进行的克隆,保真度 F 为小于 1 的恒定值,这种克隆称为普适量子克隆.

概率量子克隆(probabilistic quantum cloning) 量子力学中的一个定理. 量子不可克隆定理禁止对完全未知量子态的确定性克隆,但并未禁止只以一定概率成功的克隆,这种克隆称为概率量子克隆,其定理如下:对于给定量子态集合 $\{|\Psi_1>,|\Psi_2>,\cdots,|\Psi_n>\}$ 中的任一量子态 $|\Psi_i>$,当且仅当 $|\Psi_1>,|\Psi_2>,\cdots,|\Psi_n>$ 线性无关时,$|\Psi_i>$ 可以通过幺正-坍缩过程以大于零的成功概率精确克隆,其克隆效率依赖于态集,对于非正交态集而言,克隆效率总是小于 1.

数字信号处理

数字信号处理(digital signal processing) 一种信息处理技术. 它是 20 世纪 60 年代随着数字技术的发展而发展起来的一门新的学科领域. 它是利用计算机或专用数字处理设备,采用数值计算的方法,对信号进行加工处理,把一个输入信号变换成为另一个信号输出,以达到预期目的. 数字信号处理技术,由于它具有一系列优点,所以应用广泛,发展非常迅速. 其主要内容包括信号分析、滤波、功率谱估计、自适应信号处理等技术及其软件实现和硬件实现. 主要应用领域有通信系统、语音处理、图象处理、模式识别、人工智能、生物医学、地质勘探、气象预报、自动控制、遥感遥测、电子对抗等各个工程领域. 数字信号处理的发展速度是惊人的,它随着计算机技术的发展和高速数字信号处理器的面世而日新月

异. 现在已经发展成为一门独立的学科，它的应用几乎已经深入到人类社会生活的各个方面.

线性系统(linear system)　一种信号变换系统. 输入信号与输出信号之间满足叠加原理的系统称为线性系统. 设系统输入与输出信号之间的变换关系为 $T[\cdot]$，系统输入 $x_1(n)$ 时，对应的输出响应为 $y_1(n)$；系统输入 $x_2(n)$ 时，对应的输出响应为 $y_2(n)$，则当且仅当

$$\begin{aligned} &T[ax_1(n)+bx_2(n)] \\ &= aT[x_1(n)]+bT[x_2(n)] \\ &= ay_1(n)+by_2(n) \end{aligned}$$

时，该系统称为线性系统. 数字信号处理所讨论的系统，大多是稳定的、因果非移变的、离散线性系统.

双线性变换(bilinear transformation)　无限单位脉冲响应(IIR)数字滤波器的一种有效的设计方法. 它就是将模拟滤波器的传输函数 $H(s)$ 通过双线性变换关系

$$s = C\frac{1-z^{-1}}{1+z^{-1}} \quad (C\text{ 是待定常数}),$$

直接转换成 IIR 数字滤波器的系统函数 $H(z)$. 双线性变换是一种非线性映射关系. 由于它能将无限的模拟频率范围压缩到有限频率范围，避免从连续域变换到离散域引起的频谱混叠，因此，应用范围较广，可用来设计各种幅频特性的 IIR 数字滤波器. 该设计方法会引入频率非线性失真，必须通过频率预畸变措施，把握住关键频率点，使设计达到满意效果.

稳定系统(stable system)　简称 BIBO 系统. 一种信号变换系统. 当输入有界时，输出也有界，这种系统称为稳定系统. 线性非移变系统具有稳定性的充分必要条件是它的单位脉冲响应 $h(n)$ 绝对可和，即

$$\sum_{k=-\infty}^{+\infty} |h(k)| < +\infty.$$

数字信号处理所讨论的系统通常都是稳定的线性非移变系统.

因果系统(causal system)　一种信号变换系统. 输出的变化一定发生在输入变化之后，这种系统称为因果系统. 线性非移变系统具有因果性的充分必要条件是：当 $n<0$ 时，系统的单位脉冲响应 $h(n)=0$. 数字信号处理所讨论的实际系统，通常都具有因果性.

非移变系统(shift-invariant system)　一种信号变换系统. 系统的输入与输出特性不随时间(独立变量)的起点变化而变化，即：如果 $y(n)$ 是系统对于 $x(n)$ 的响应，则 $y(n-k)$ 就是该系统对于 $x(n-k)$ 的响应(k 是任意整数，表示位移). 此类系统称为非移变系统. 当 n 表示离散时间变量，则此类系统称为非时变系统.

非时变系统(time-invariant system)　见“非移变系统”.

线性非移变系统(linear shift-invariant system)　一种信号变换系统. 同时具有线性和非移变特性的系统，称为线性非移变系统. 这种系统的变换特性可以用它的单位脉冲响应来表征.

线性非移变系统级联(cascade combination of linear shift-invariant system)　系统之间的一种连结方式. 两个线性非移变系统 $h_1(n)$ 和 $h_2(n)$，如果按照图1所示方式连结，则称为线性非移变系统级



图1

联. 级联之后，仍构成一个线性非移变系统，其单位脉冲响应 $h(n)$ 等于 $h_1(n)$ 和 $h_2(n)$ 的卷积和，即等效于图 2 所示形式.

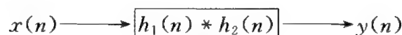


图2

线性非移变系统并联(parallel combination of linear shift-invariant system)　系统之间的一种连结方式. 两个线性非移变离散系统 $h_1(n)$ 和 $h_2(n)$，

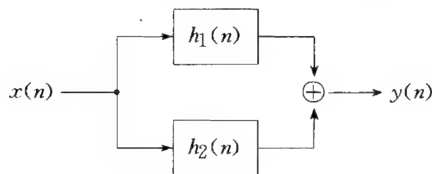


图1

如果按照图 1 所示方式连结，则称为线性非移变离散系统的并联. 并联之后仍构成一个线性非移变离散系统，其单位脉冲响应 $h(n)$ 等于 $h_1(n)$ 和 $h_2(n)$ 的代数和. 即等效于图 2 所示形式.

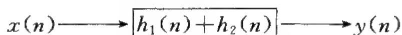


图2

系统函数(system function)　描述离散系统的一个函数. 当线性非移变离散系统初始不储能情况下，其输出序列的 Z 变换与输入序列的 Z 变换之比，称为该系统的系统函数，即

$$H(Z)=\frac{Y(Z)}{X(Z)},$$

式中，$X(Z)$ 是系统输入 $x(n)$ 的 Z 变换，$Y(Z)$ 是系统输出 $y(n)$ 的 Z 变换. 实质上，离散系统的系统函数 $H(Z)$ 就是系统单位脉冲响应 $h(n)$ 的 Z 变换. 如同 $h(n)$，$H(Z)$ 能用来表征一个离散系统.

离散系统(discrete system)　一种信号系统. 当一个系统的输入是离散信号，输出也是离散信号，则该系统称为离散系统.

数字系统(digital system) 一种信号系统. 当一个系统的输入是数字信号,输出也是数字信号,则该系统称为数字系统. 换句话说,数字系统用来处理数字信号,它是由数字器件构成的.

卷积和(convolution sum) 一种数学运算方法. 序列 $x_1(n)$和 $x_2(n)$进行如下形式的运算:

$$\sum_{k=-\infty}^{\infty} x_1(k) \cdot x_2(n-k),$$

则称为卷积和. 又称为线性卷积,通常记为

$$x_1(n) * x_2(n).$$

卷积和运算是离散线性系统固有的一种基本运算,它符合交换律,即

$$x_1(n) * x_2(n) = x_2(n) * x_1(n).$$

线性卷积(linear convolution) 即"卷积和". 又称离散卷积.

周期卷积(periodic convolution) 一种数学运算方法. 周期长度均为 N 的两个周期序列 $\tilde{x}_1(n)$和 $\tilde{x}_2(n)$进行如下形式的运算:

$$\sum_{m=0}^{N-1} \tilde{x}_1(m) \cdot \tilde{x}_2(n-m),$$

称为周期卷积. 通常记为 $\tilde{x}_1(n) \circledast \tilde{x}_2(n)$. 周期卷积的结果仍然是以 N 为周期的序列,其运算符合交换律,即

$$\tilde{x}_1(n) \circledast \tilde{x}_2(n) = \tilde{x}_2(n) \circledast \tilde{x}_1(n).$$

离散傅氏变换的循环卷积(discrete Fourier transform circular convolution) 一种数学运算方法. 长度均为 N 的两个有限长序列 $x_1(n)$ 和 $x_2(n)$ 进行如下形式的运算:

$$\left[\sum_{m=0}^{N-1} x_1((m))_N \cdot x_2((m))_N\right] \cdot R_N(n),$$

式中 $x_1((m))_N$ 表示 $x_1(m)$ 以 N 为周期进行周期延拓所构成的周期序列;$x_2((m))_N$ 表示 $x_2(-m)$ 以 N 为周期进行周期延拓后作线性移位,即 $x_2(m)$ 进行循环移位所构成的周期序列;

$$R_N(n) = \begin{cases} 1 & (n = 0 \sim N-1), \\ 0 & (\text{其他}) \end{cases}$$

称为循环卷积,通常记为 $x_1(n) Ⓝ x_2(n)$. 循环卷积等于周期卷积取主值序列,即

$$x_1(n) Ⓝ x_2(n) = \tilde{x}_1(n) \circledast \tilde{x}_2(n) \quad (n = 0 \sim N-1).$$

循环卷积运算符合交换律,即

$$x_1(n) Ⓝ x_2(n) = x_2(n) Ⓝ x_1(n).$$

二维卷积和(two-dimensional convolution sum) 一种数学运算方法. 两个二维序列 $x_1(m,n)$ 和 $x_2(m,n)$ 进行如下形式的运算:

$$\sum_{k=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} x_1(k,r) \cdot x_2(m-k, n-r),$$

称为二维卷积和,记为 $x_1(m,n) * x_2(m,n)$. 二维卷积和运算符合交换律,即

$$x_1(m,n) * x_2(m,n) = x_2(m,n) * x_1(m,n).$$

取样率(sampling rate or sampling frequency) 数字信号处理的一个基本概念. 对连续时间信号进行取样时,如果采用的取样间隔时间为 T,则 T 的倒数称为取样速率或取样频率,简称取样率. 实际上,它表示单位时间抽取的样点数.

取样定理(sampling theorem) 数字信号处理的基本定理之一. 对于一个具有有限频带宽度的模拟信号 $x(t)$进行均匀取样,当取样率 f_s 大于或等于 $x(t)$的频谱中最高频率 f_m 的两倍时,则能够从样点集合准确地恢复原信号 $x(t)$. 因为这时从样点所得到的离散信号 $x(k)$,它的频谱是以 $x(t)$的频带宽度为周期的周期性频谱,包含有原信号的全部信息. 即

$$X_s(\Omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X(\Omega - k\Omega_s),$$

式中 $X(\Omega)$为 $x(t)$的频谱;$X_s(\Omega)$为 $x(k)$的频谱;$X(\Omega)$的最高角频率为 $\Omega_m = 2\pi f_m$,取样率为 $f_s \geqslant 2f_m$;$X_s(\Omega)$的周期为 $\Omega_s = 2\pi f_s$. 取样定理是实现信号的数字处理代替模拟处理的理论依据,对研究信息传输和信号的数字处理有着非常重要的意义,在实际中获得广泛应用.

循环卷积的离散傅里叶变换(discrete Fourier transform of circular convolution) 一种数字变换方法. 设循环卷积 $x_3(n) = x_1(n) Ⓝ x_2(n)$,则有

$$X_3(k) = X_1(k) \cdot X_2(k),$$

式中 $X_1(k), X_2(k), X_3(k)$ 分别是 $x_1(n), x_2(n), x_3(n)$ 的 N 点离散傅里叶变换.

Z 变换(Z-transform) 函数的一种变换方法. 序列 $x(n)$ 的 Z 变换定义为

$$X(Z) = \sum_{n=-\infty}^{\infty} x(n) Z^{-n},$$

式中 Z 为复变量,$X(Z)$ 为复变函数 .

二维Z 变换(two-dimensional Z-transform) 一种数字变换方法. 二维序列 $x(m,n)$ 的二维 Z 变换 $X(Z_1,Z_2)$ 定义为

$$X(Z_1,Z_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(m,n) \cdot Z_1^{-m} Z_2^{-n},$$

式中 Z_1, Z_2 都是复变量,$X(Z_1,Z_2)$ 是二维复变函数.

二维卷积定理(two-dimensional convolution theorem) 数字信号处理的基本定理之一. 对于二维线性非移变系统,输出序列 $y(m,n)$等于输入序列 $x(m,n)$与单位脉冲响应序列 $h(m,n)$的卷积和,即

$$y(m,n) = x(m,n) * h(m,n)$$

$$= \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} x(n_1,n_2)h(m-n_1,n-n_2)$$

$$= \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} h(n_1,n_2)x(m-n_1,n-n_2).$$

离散傅里叶级数(discrete Fourier series) 简称 DFS. 一种数字变换. 将周期序列展开为傅里叶级数,称为离散傅里叶级数. 其变换式可写为

$$\begin{cases} \widetilde{X}(k) = \sum_{n=0}^{N-1} \widetilde{x}(n) \cdot W_N^{kn}, \\ \widetilde{x}(n) = \frac{1}{N}\sum_{n=0}^{N-1} \widetilde{X}(k) \cdot W_N^{-kn}, \end{cases}$$

式中 $\widetilde{x}(n)$为时域周期序列;$\widetilde{X}(k)$为频域周期序列;

$$W_N = e^{-j2\pi/N}.$$

离散傅里叶变换(discrete Fourier transform) 简称 DFT. 一种数字变换方法. 时域有限序列$x(n)$的离散傅里叶变换 $X(k)$定义为

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot W_N^{kn},$$

式中 $k=0\sim N-1$,$W_N=e^{-j2\pi/N}$,时域序列 $x(n)$的长度为 N,频域序列 $X(k)$的长度为 N. 上式通常称为 N 点离散傅里叶变换.

离散傅里叶反变换(inverse discrete Fourier transform) 简称 IDFT. 一种数字变换的逆变换. 频域有限序列 $X(k)$ 的离散傅里叶反变换 $x(n)$ 定义为

$$x(n) = \frac{1}{N}\sum_{n=0}^{N-1} X(k) \cdot W_N^{-kn},$$

式中 $n=0\sim N-1$,$W_N=e^{-j2\pi/N}$,时域序列 $x(n)$ 的长度为 N,频域序列 $X(k)$ 的长度为 N.

二维离散傅里叶变换(two-dimensional discrete Fourier transform) 一种数字变换方法. 在空间域一个有限长度二维序列 $x(m,n)$ 的二维离散傅里叶变换 $\widetilde{X}(k,l)$ 定义为

$$\widetilde{X}(k,l) = \left[\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} x(m,n)W_M^{km}W_N^{ln}\right] \cdot R_{MN}(k,l),$$

式中 $x(m,n)$的区域范围为:

$m=0\sim M-1$;

$n=0\sim N-1$;

$W_M=e^{-j2\pi/M}$;

$W_N=e^{-j2\pi/N}$;

$$R_{MN}(k,l)=\begin{cases}1 & (k=0\sim M-1, l=0\sim N-1), \\ 0 & (\text{其他}).\end{cases}$$

二维离散傅里叶反变换(two-dimensional inverse discrete Fourier transform) 一种数字变换的逆变换. 在频域一个有限长度二维序列 $X(k,l)$ 的二维离散傅里叶反变换 $x(m,n)$ 定义为

$$x(m,n) = \frac{1}{MN}\left[\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} X(k,l)W_M^{-km}W_N^{-ln}\right] \cdot R_{MN}(m,n),$$

式中 $X(k,l)$的区域范围为:

$k=0\sim M-1$;

$l=0\sim N-1$;

$W_M=e^{-j2\pi/M}$;

$W_N=e^{-j2\pi/N}$;

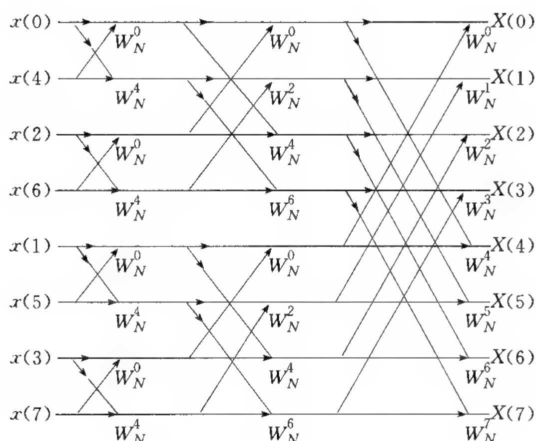
$$R_{MN}(m,n)=\begin{cases}1 & (m=0\sim M-1, n=0\sim N-1), \\ 0 & (\text{其他}).\end{cases}$$

频域取样定理(sampling theorem in the frequency domain) 数字信号处理的基本定理之一. 对于有限时宽序列 $x(n)$的周期连续频谱 $X(e^{j\omega})$进行均匀取样,当一个周期内的取样点数 N 大于或至少等于 $x(n)$的有限时宽时,则有可能从频谱样点 $X(k)$中无失真地恢复原来的周期连续频谱. 频域取样定理之所以重要,在于它揭示了连续周期频谱与离散周期频谱之间的内在联系. 如果已知一个信号的频谱,只要符合频域取样定理,对它进行频率取样,则有可能利用数字的方法求得相应的信号,从而为数字信号处理技术开拓了新的途径.

快速傅里叶变换(fast Fourier transform) 简称 FFT. 一类快速算法的总称. 离散傅里叶变换由于计算量大,实时性差,在实际应用中受到很大限制. 1965 年提出快速算法,使离散傅里叶变换的复数乘法计算量从 N^2 次减少到$(N\log_2 N)/2$ 次($N=2^r$,r 为正整数),使运算速度大为提高. 此后,又导致了若干快速傅里叶变换算法的出现. 因此,人们把以库利-图基算法为基础的这类算法统称为快速傅里叶变换. 快速傅里叶变换是数字频谱分析的基础,在数字信号处理学科发展过程中起着重要作用,在各学科领域得到广泛应用. 快速傅里叶变换的基本原理是:把计算长度为 N 的时间序列的离散傅里叶变换,逐次分解为计算长度较短的序列的离散傅里叶变换,充分利用变换核 W_N^{kn} 的周期性和对称性,从而使计算量大大减少,计算效率大大提高.

时间抽选快速傅里叶变换算法(decimation-in-time fast Fourier transform algorithm) 一种计算离散傅里叶变换的快速算法. 这类算法的基本思路是:将离散傅里叶变换(DFT)的输入序列在时域进行抽选(参见"离散傅里叶变换"),逐次分解为越来越短的子序列,然后分别进行较短序列的离散傅里叶变换,最后再按照一定规律逐次组合,构成原序列的离散傅里叶变换. 以序列长度 N 为 2 的整幂次基 2FFT 为例(参见"快速傅里叶变换"). 设 $x(n)$ 为 N 点时间序列,$N=2^v$,v 为正整数. $x(n)$ 的 DFT 为

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{nk} \quad (k=0\sim N-1),$$



将 $x(n)$ 分解成偶数点和奇数点子序列，

$$X(k)=\sum_{r=0}^{\frac{N}{2}-1}x(2r)W_{N/2}^{rk}+W_N^k\sum_{r=0}^{\frac{N}{2}-1}x(2r+1)W_{N/2}^{rk}$$
$$=G(k)+W_N^kH(k),$$

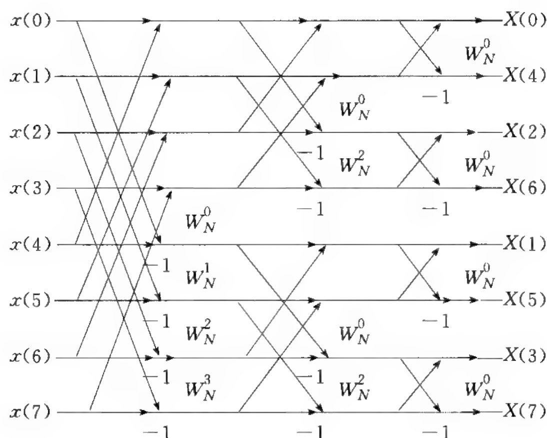
式中 $G(k)$ 和 $H(k)$ 分别是分解成的偶数点和奇数点序列的 $N/2$ 点 DFT. 按照上述原理，可将 $G(k)$ 和 $H(k)$ 的输入序列分别在时域继续分解成 4 组 $N/4$ 点的 DFT. 以此类推，直至最后分解成一系列 2 点的 DFT，便可得到所谓按时间抽选的 FFT 算法. 若 $N=8$，则可得到 8 点按时间抽选的 FFT 算法，流图如图所示.

频率抽选快速傅里叶变换算法(decimation-in-frequency fast Fourier transform algorithm)　一种计算离散傅里叶变换的快速算法. 这类算法的基本思路是：将离散傅里叶变换(DFT)的输入序列在频域进行抽选(参见“离散傅里叶变换”)，逐次分解成越来越短的子序列，然后将其所对应的较短的时间序列分别进行离散傅里叶变换，最后再按照一定的规律逐次组合，构成原序列的离散傅里叶变换. 以序列的长度 N 为 2 的整幂次基 2 FFT 为例(参见“快速傅里叶变换”). 设 $x(n)$ 为 N 点时间序列，$N=2^v$，v 为正整数. $x(n)$ 的 DFT 为 $X(k)$，是 N 点频序列. 将 $X(k)$ 分解成偶数点和奇数点子序列，则

$$X(2r)=\sum_{n=0}^{N/2-1}\left[x(n)+x\left(n+\frac{N}{2}\right)\right]\cdot W_{N/2}^{rn}$$
$$=\sum_{n=0}^{N/2-1}g(n)\cdot W_{N/2}^{rn},$$

$$X(2r+1)$$
$$=\sum_{n=0}^{N/2-1}\left[x(n)-x\left(n+\frac{N}{2}\right)\right]\cdot W_N^n\cdot W_{N/2}^{rn}$$
$$=\sum_{n=0}^{N/2-1}h(n)\cdot W_N^n\cdot W_{N/2}^{rn},$$

式中 $r=0\sim N/2-1$. 这类算法是先形成序列 $g(n)$ 和 $h(n)$，再计算 $h(n)\cdot W_N^n$，然后分别计算这两个子序列的 $N/2$ 点 DFT，最后得到偶数点 DFT 和奇数点 DFT 的输入. 按照上述原理，继续在频域将输出序列分解成更短序列，直至不能再分解，便可得到所谓按频率抽选的 FFT 算法. 若 $N=8$，则可得到 8 点按频率抽选的 FFT 算法，流图如图所示.



同址运算(computation in place)　编制快速傅里叶变换(FFT)计算程序的具体编程技巧问题(参见“快速傅里叶变换”). 所谓同址运算，即指在同一组计算机内存空间地址上进行计算. 根据 FFT 算法流图的结构特点，其输入数据、各级中间数据、输出数据，不仅个数相同，而且位置相互对应，上一级蝶形运算输出的数据作为下一级蝶形运算的输入，因此，可利用同一组内存地址进行计算. 由于 FFT 算法可以实现同址运算，因而能够节省大量运算单元，具有重要意义.

数字滤波器(digital filter)　一个数字处理系统. 它通过对输入信号序列进行特定的数字运算，得到另一输出序列，从而达到抑制干扰、提取有用信号的效果. 数字滤波器通常要根据其单位脉冲响应的长度划分为两类. 一类，单位脉冲响应有限长，硬件实现的结构没有反馈环节，称为有限单位脉冲响应数字滤波器，或非递归型数字滤波器，简称 FIR 数字滤波器. 另一类，单位脉冲响应无限长，硬件实现的结构具有反馈环节，称为无限单位脉冲响应数字滤波器，或递归型数字滤波器，简称 IIR 数字滤波器. 数字滤波器的设计与实现，是数字信号处理的重要内容，在各工程领域得到广泛应用.

相关序列和协方差序列的性质(property of correlation and covariance sequence)　一种特定含义的数学属性. 对随机信号进行处理时，经常用到相关序列和协方差序列的性质. 设 $\{x_n\}$ 和 $\{y_n\}$ 为两个实平稳随机过程. 自相关为 $\varphi_{xx}(m)=E[x_nx_{n+m}]$，自协方差为

$$r_{xx}(m)=E[(x_n-m_x)(x_{n+m}-m_x)],$$

互相关为 $\varphi_{xy}(m)=E[x_ny_{n+m}]$，互协方差为

$$r_{xy}(m)=E[(x_n-m_x)(y_{n+m}-m_y)],$$

其中 m_x 是 $\{x_n\}$ 的均值，m_y 是 $\{y_n\}$ 的均值. 它们有以下性质：

1. $r_{xx}(m)=\varphi_{xx}(m)-m_x^2$,

$r_{xy}(m)=\varphi_{xy}(m)-m_x m_y$,

当 $m_x=0$ 时，$r_{xx}(m)=\varphi_{xx}(m)$，$r_{xy}(m)=\varphi_{xy}(m)$.

2. $\varphi_{xx}(0)=E[x_n^2]=$均方值,

$r_{xx}(0)=\sigma_x^2=$方差.

3. $\varphi_{xx}(m)=\varphi_{xx}(-m)$；$r_{xx}(m)=r_{xx}(-m)$,

$\varphi_{xy}(m)=\varphi_{yx}(m)$；$r_{xy}(m)=r_{yx}(-m)$.

4. $|\varphi_{xy}(m)|\leqslant\sqrt{\varphi_{xx}(0)\varphi_{yy}(0)}$,

$|r_{xy}(m)|\leqslant\sqrt{r_{xx}(0)r_{yy}(0)}$,

$|\varphi_{xx}(m)|\leqslant\varphi_{xx}(0)$,

$|r_{xx}(m)|\leqslant r_{xx}(0)$.

5. 如果 $y_n=x_{n-n_0}$，则

$\varphi_{yy}(m)=\varphi_{xx}(m)$，　$r_{yy}(m)=r_{xx}(m)$.

6. 对于多数随机过程，随机变量之间的相关性将随着时间的延长而减小，因此

$$\lim_{m\to\infty}\varphi_{xx}(m)=m_x^2,\qquad \lim_{m\to\infty}r_{xx}(m)=0.$$

$$\lim_{m\to\infty}\varphi_{xy}(m)=m_x m_y,\qquad \lim_{m\to\infty}r_{xy}(m)=0.$$

这说明相关序列和协方差序列是非周期序列，它们将随 m 值的增大而逐渐消失，所以，它们的 Z 变换是存在的.

功率谱(power spectrum)　平稳随机信号的功率在各个不同的频率上的分布规律的描述. 根据随机过程平均的概念，当随机变量 x_n 的均值为零时，其方差等于平均功率，因此

$$\sigma_x^2=\frac{1}{2\pi}\int_{-\pi}^{\pi}P_{xx}(\omega)\mathrm{d}\omega.$$

函数 $P_{xx}(\omega)$ 在某一频带上的积分，正比于信号在该频带上的功率，所以，人们称 $P_{xx}(\omega)$ 为功率密度谱，简称功率谱. 在均值为零的情况下，信号的功率谱是其自相关序列的傅里叶变换. 但是，在均值不为零的情况下，其傅里叶变换不存在. 为了避免这一困难，在信号处理领域，定义功率谱为协方差序列的傅里叶变换，即

$$P_{xx}(\omega)=\Gamma_{xx}(\mathrm{e}^{\mathrm{j}\omega}),$$

其中 $\Gamma_{xx}(\mathrm{e}^{\mathrm{j}\omega})$ 为 $r_{xx}(m)$ 的傅里叶变换.

功率谱分为自功率谱 $P_{xx}(\omega)$ 与互功率谱 $P_{xy}(\omega)$，是随机信号特征在频域的描述. 同样，互功率谱定义为

$$P_{xy}(\omega)=\Gamma_{xy}(\mathrm{e}^{\mathrm{j}\omega}),$$

其中 $\Gamma_{xy}(\mathrm{e}^{\mathrm{j}\omega})$ 为 $r_{xy}(m)$ 的傅里叶变换.

线性系统对随机信号的响应(response of linear system to random signal)　线性系统分析中的基本概念. 对于稳定的线性非移变系统，如果输入随机过程是平稳的，则输出随机过程也是平稳的. 如果输入随机过程的均值为 m_x，则输出随机过程的均值为

$$m_y=E[y(n)]=m_x\sum_{k=-\infty}^{\infty}h(k),$$

其中 $h(k)$ 为线性系统的单位脉冲响应. 输出自相关序列，等于输入自相关序列与系统的单位脉冲响应自相关序列的卷积和，即

$$\begin{aligned}\varphi_{yy}(m)&=\sum_{l=-\infty}^{\infty}\varphi_{xx}(m-l)v(l)\\&=h(m)*h(-m)*\varphi_{xx}(m),\end{aligned}$$

其中

$$v(l)=\sum_{k=-\infty}^{\infty}h(k)h(k+l).$$

输入与输出之间的互相关序列，等于系统单位脉冲响应与输入自相关序列的卷积和，即

$$\begin{aligned}\varphi_{xy}(m)&=\sum_{k=-\infty}^{\infty}h(k)\cdot\varphi_{xx}(m-k)\\&=h(m)*\varphi_{xx}(m).\end{aligned}$$

当输入为白噪声时，即

$$m_x=0;\ \varphi_{xx}(m)=\sigma_x^2\delta(m),$$

$$P_{xx}(\omega)=\sigma_x^2\quad(-\pi\leqslant\omega\leqslant\pi),$$

则输入与输出的互相关为

$$\varphi_{xy}(m)=\sigma_x^2h(m);$$

输入与输出的互功率谱为

$$P_{xy}(\omega)=\sigma_x^2H(\mathrm{e}^{\mathrm{j}\omega}).$$

FIR 数字滤波器(FIR digital filter)　有限单位脉冲响应数字滤波器的简称. 见“数字滤波器”.

IIR 数字滤波器(IIR digital filter)　无限单位脉冲响应数字滤波器的简称. 见“数字滤波器”.

脉冲响应不变法(impulse invariant method)　无限单位脉冲响应(IIR)数字滤波器的一种设计方法. 它是将模拟滤波器的冲激响应 $h(t)$ 直接进行取样，而作为 IIR 数字滤波器的单位脉冲响应 $h(n)$，即

$$h(n)=h(t)|_{t=nT}.$$

IIR 数字滤波器的脉冲响应不变法设计是建立在取样定理基础上的，只有在满足取样定理时才能加以应用. 也就是说，脉冲响应不变法通常只能用来设计具有频带有限的数字滤波器.

误差准则(error criterion)　网络设计的一种标准. 在各种网络综合的优化设计中，为了实现最佳逼近，通常要根据一定的误差准则，确定一个满足给定误差准则条件下的网络函数，从而达到优化设计的目的. 这种在最佳逼近过程中根据技术指标误差的要求，使误差最小所采用的准则，称为误差准则. 常用的误差准则有最小平方误差准则(最小二乘法)、均方误差最小化准则、P 次方误差最小化准则和最大绝对误差最小化准则等.

均方误差最小化准则(minimization criterion of mean-square error)　一种优化设计的误差准则.就是使未知量与已知量的均方误差达到最小化,在这种条件下来确定所需的未知量.这个准则可应用于滤波器优化设计.如在频域设计一个无限单位脉冲响应(IIR)滤波器.假设滤波器所要求的频率特性为 $H_d(e^{j\omega})$,待设计滤波器的系统函数为

$$H(Z)=A\prod_{k=1}^{N}\frac{1+a_kZ^{-1}+b_kZ^{-2}}{1+c_kZ^{-1}+d_kZ^{-2}},$$

相应的频率特性可写成为

$$H(e^{j\omega})=H(Z)\,|_{Z=e^{j\omega}},$$

则在一组离散频率点上 $\{\omega_i\}$ $(i=1,2,\cdots,M)$所定义的均方误差函数为

$$E=\sum_{i=1}^{M}[\,|H(e^{j\omega})H(e^{j\omega})-H_d(e^{j\omega})|\,]^2,$$

然后通过优化算法,求出滤波器参数,使均方误差 E 达到最小.

P 次方误差最小化准则(minimization of a P-error criterion)　一种优化设计的误差准则.就是使未知量与已知量的 P 次方加权平均误差达到最小化,在这种条件下来确定所需的未知量.这个准则若在频域应用于无限单位脉冲响应滤波器设计,则根据频域 P 次方加权平均误差最小化的要求达到优化设计的目的.其设计方法如下:设 $H_d(e^{j\omega})$为所要求的频率特性,$H(e^{j\omega})$为待设计滤波器的频域特性,则 P 次方加权平均误差函数可定义为

$$E_P=\int_0^{\pi}W(\omega)[\,|H(e^{j\omega})-H_d(e^{j\omega})|\,]^P\mathrm{d}\omega,$$

或

$$E'_P=\int_0^{\pi}W(\omega)[\tau(\omega)-\tau_d(\omega)]^P\mathrm{d}\omega,$$

其中 $W(\omega)$为加权函数,E_P 为幅度 P 次方加权平均误差,E'_P 为相位 P 次方加权平均误差.然后通过优化算法,使 E_P 或 E'_P 达到最小,从而求得滤波器的有关参数.

最小平方逆设计(least-square inverse design)　无限单位脉冲响应(IIR)滤波器的一种计算机辅助优化设计方法.这种设计方法的原理是对所求滤波器的逆滤波器作最小平方逼近.设理想的单位脉冲响应前 L 个值为

$$\{h_d(n)\}\quad(n=0\sim L-1),$$

IIR 滤波器的系统函数为

$$H(Z)=\frac{b_0}{1-\sum_{k=1}^{N}a_kZ^{-k}},$$

则当输入为 $h_d(n)$ 时,逆滤波器 $1/H(Z)$ 的输出 $v(n)$ 应逼近单位取样脉冲.由

$$\frac{1}{H(Z)}=\frac{v(Z)}{H_d(Z)}$$

得到

$$b_0v(n)=h_d(n)-\sum_{r=1}^{N}a_rh_d(n-r),$$

令 $b_0=h_d(0)$,且 $n>0$ 时 $v(n)$ 应尽可能小,即

$$\begin{aligned}E&=\sum_{n=1}^{\infty}[v(n)]^2\\&=\frac{1}{b_0^2}\sum_{n=1}^{\infty}[h_d(n)]^2-2\sum_{n=1}^{\infty}h_d(n)\sum_{r=1}^{N}a_r\,h_d(n-r)\\&\quad+\sum_{n=1}^{\infty}[\sum_{r=1}^{N}a_rh_d(n-r)]^2\end{aligned}$$

应达到最小.令

$$\frac{\partial E}{\partial a_i}=0\quad(i=1\sim N),$$

由此得到线性方程组,解线性方程组,即可确定 IIR 滤波器的系数 a_k.这就是所谓 IIR 滤波器的最小平方逆设计.

窗函数(window function)　一种特殊函数.指具有有限长度的、形如在时域或频域开一个窗口的序列函数.常用的窗函数有海宁窗、哈明窗、凯塞窗等各种类型.窗函数起着截短和平滑的作用,可用来设计滤波器或进行频谱分析.由于不同的窗函数频谱特性不同,可根据需要选用不同的窗函数.例如,海宁窗(升余弦窗)函数表达式为

$$W_H(n)=\frac{1}{2}\left(1-\cos\frac{2\pi n}{N-1}\right)\quad(0\leqslant n\leqslant N-1);$$

布拉克曼窗函数表达式为

$$W_B(n)=0.42-0.5\cos\frac{2\pi n}{N-1}+0.08\cos\frac{4\pi n}{N-1}\quad(0\leqslant n\leqslant N-1).$$

FIR 滤波器的窗函数设计法(design of FIR filters using window function)　有限单位脉冲响应(FIR)滤波器的一种设计方法.利用窗函数进行 FIR 滤波器设计,这是一种最直接的设计方法.它的设计思想是:把求得的无限单位脉冲响应,通过窗函数直接截取为有限序列,作为 FIR 滤波器的单位脉冲响应.设 $H_d(e^{j\omega})$ 为指标所要求的滤波器的频率响应,其对应的单位脉冲响应为无限长,可表示为

$$h_d(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}H_d(e^{j\omega})\cdot e^{j\omega}\cdot\mathrm{d}\omega.$$

若采用的窗函数为 $\omega(n)$,$0\leqslant n\leqslant N-1$,则所设计的 FIR 滤波器的单位脉冲响应为

$$h(n)=h_d(n)\cdot\omega(n).$$

切比雪夫窗(Chebyshev window)　一种局部优化的时窗函数.它满足窗函数的最大振幅比准则,所以也称为最大振幅比时窗函数.用切比雪夫窗来设计 FIR 滤波器,比采用通常的时窗函数具有明显的优点.切比雪夫窗的数学定义如下:

时域表示为

$$w(n)=\frac{\sum_{n=-M}^{M}W(m)\cos\left(\frac{2\pi}{N}mn\right)}{\sum_{n=-M}^{M}W(m)}\quad(|n|\leqslant M),$$

离散频谱

$$W(m)=R_P\cos\left[M\mathrm{arc}\cos\frac{2\cos\left(\frac{2\pi m}{N}\right)-\cos(2\pi\cdot\Delta F)+1}{\cos(2\pi\cdot\Delta F)+1}\right],$$

式中参数 $N,R_P,\Delta F,M$ 之间的相对关系为

阶数 $N=2M+1$ 或 $M=(N-1)/2$，

$$N\geqslant 1+\frac{\mathrm{ch}\left[\frac{1+R_P}{R_P}\right]}{\mathrm{ch}\left[\frac{1}{\cos(\pi\cdot\Delta F)}\right]},$$

$$\Delta F=\frac{1}{\pi}\arccos\left\{\frac{1}{\mathrm{ch}\left[\frac{\mathrm{ch}\left(\frac{1+R_P}{R_P}\right)}{N-1}\right]}\right\},$$

$$R_P=\frac{1}{\mathrm{ch}\left[(N-1)\arccos\left(\frac{1}{\cos(\pi\cdot\Delta F)}\right)\right]}-1.$$

凯塞窗(Kaiser window)　一种局部优化的具有较强能力的窗函数.它是凯塞(Kaiser)采用修正的零阶贝塞尔函数实现的.已经证明,凯塞窗是十分接近最大能量比的窗函数.用它来设计有限单位脉冲响应(FIR)滤波器可以有效地调整窗函数频谱主瓣与旁瓣的相对关系,从而使设计的 FIR 滤波器能够同时满足通带波动、阻带衰耗和过渡带宽等多项设计指标的要求.凯塞窗函数的表示式为

$$w(n)=\begin{cases}\frac{I_0(\alpha)}{I_0(\beta)} & \left(|n|\leqslant\frac{N-1}{2}\right),\\ 0 & \left(|n|>\frac{N-1}{2}\right),\end{cases}$$

其中

$$\alpha=\beta\sqrt{1-\left(\frac{2n}{N-1}\right)^2},$$

$$I_0(x)=1+\sum_{k=1}^{\infty}\left[\frac{(x/2)^k}{k!}\right]^2,$$

$$\beta=\begin{cases}0.1102(A_a-8.7) & (A_a>50\ \mathrm{dB}),\\ 0.58417(A_a-20.96)^{0.4}\\ +0.07886(A_a-20.96)\\ (20.96\mathrm{dB}\leqslant A_a\leqslant 50\ \mathrm{dB}),\\ 0.0 \quad (A_a<20.96\ \mathrm{dB}).\end{cases}$$

$$N=\frac{D}{\Delta F}+1,$$

$$D=\begin{cases}\frac{A_a-7.95}{14.36} & (A_a>20.96\mathrm{dB}),\\ 0.9222 & (A_a\leqslant 20.96\mathrm{dB}),\end{cases}$$

A_a 为阻带衰耗,单位是 dB;ΔF 为归一化过渡带宽.

最大能量比(largest energy ratio)　一种优化设计方法.可作为窗函数优化的准则之一.在给定窗函数频谱主瓣宽度的条件下,其主瓣能量与频谱总能量之比,称为窗函数的能量比.即设窗函数为 $w(n)$，$|n|\leqslant M$，$N=2M+1$.其频谱为

$$W(f)=\sum_{n=-M}^{M}w(n)\mathrm{e}^{-\mathrm{j}2\pi fn}\quad\left(|f|\leqslant\frac{f_s}{2}\right).$$

给定主瓣宽度$[-\delta,\delta]$ $(0<\delta<f_s/2)$,则能量比定义为

$$T[w(n)]=\frac{\int_{-\delta}^{\delta}|W(f)|^2\mathrm{d}f}{\int_{-\frac{1}{2}f_s}^{\frac{1}{2}f_s}|W(f)|^2\mathrm{d}f}.$$

所谓最大能量比,就是通过优选窗函数 $w(n)$,使能量比 $T[w(n)]$ 达到最大值.而达到最大能量比时所确定的窗函数,即最大能量比窗函数.

最大振幅比(largest amplitude ratio)　窗函数优化的准则之一.在给定窗函数频谱主瓣宽度的条件下,其主瓣幅度与旁瓣最大幅度之比,称为窗函数的振幅比.即设窗函数 $w(n)$ 为偶对称有限序列,阶数 $N=2M+1$,其频谱为

$$W(f)=\sum_{n=-M}^{M}w(n)\cos 2\pi fn\quad\left(f\leqslant\frac{f_s}{2}\right).$$

给定主瓣宽度$[-\delta,\delta]$　$(0<\delta<f_s/2)$,则振幅比定义为

$$Q[w(n)]=\frac{|W(0)|}{\max\limits_{\delta<f<\frac{1}{2}f_s}|W(f)|}=\frac{\left|\sum_{n=-M}^{M}w(n)\right|}{\max\limits_{\delta<f<\frac{1}{2}f_s}\left|\sum_{n=-M}^{M}w(n)\cos 2\pi fn\right|}.$$

所谓最大振幅比,就是通过优选窗函数 $w(n)$,使振幅比$Q[w(n)]$ 达到最大值.而达到最大振幅比时所确定的窗函数,即最大振幅比窗函数.

FIR 滤波器的线性相位特性(linear-phase property of FIR filter)　有限单位脉冲响应(FIR)滤波器的基本性质.如果 FIR 滤波器的单位脉冲响应满足以下条件:

$$h(n)=\pm h(N-1-n),$$

式中 N 为 $h(n)$ 的长度,则该 FIR 滤波器具有严格的线性相位特性.其线性相位为

$$\varphi(\omega)=-\omega\left(\frac{N-1}{2}\right)=-\omega\tau,$$

其中 τ 为延时,是常量.这是 FIR 滤波器最重要的特性,它使得 FIR 滤波器在许多要求线性相位的应用领域,如图象通信等领域得到了广泛采用.

FIR 滤波器设计的等波纹逼近法(equiripple

approximation method for FIR filter design)　有限单位脉冲响应(FIR)滤波器的一种计算机辅助的优化设计方法. 这种设计方法采用频域最大绝对误差最小化准则. 设 FIR 滤波器的幅频特性 $H(e^{j\omega})$ 与理想的幅频特性 $H_d(e^{j\omega})$ 之间的最大绝对误差为

$$\max_{\omega}|H_d(e^{j\omega}) - H(e^{j\omega})| \cdot \overline{W}(\omega),$$

式中 $\overline{W}(\omega)$ 为加权函数,则当滤波器阶数一定时,通过优化选择滤波器函数 $h(n)$,使最大绝对误差达到最小值,即

$$\min_{h(n)}\max_{\omega}|H_d(e^{j\omega}) - H(e^{j\omega})| \cdot \overline{W}(\omega),$$

这时所确定的 FIR 滤波器参数,在通带和阻带范围内,其幅频特性都具有等波动的特点. 所以,通常把这种设计方法称为 FIR 滤波器的等波纹逼近法. 关于 FIR 滤波器的等波纹逼近设计程序,最常用的是雷米兹算法. 相应的程序 1979 年被收入美国电子与电器工程协会(IEEE)数字信号处理学会编辑的《数字信号处理程序库》一书内.

交错定理(alternation theorem)　逼近理论应用于滤波器设计的一个重要定理. 它是有限单位脉冲响应(FIR)滤波器设计等波纹逼近法的理论依据. 该定理认为:令 F 为闭区间 $0 \leqslant \omega \leqslant \pi$ 的任一闭子集,为使余弦多项式

$$H(\omega) = h(0) + \sum_{n=0}^{M} 2h(n) \cdot \cos\omega n$$

在 F 上成为理想频谱 $H_d(e^{j\omega})$ 的惟一最佳逼近,其充分必要条件是绝对误差函数

$$E(\omega) = W(\omega) \cdot [H_d(e^{j\omega}) - H(\omega)]$$

在 F 上至少呈现 $(M+2)$ 个频率点,使得

$$E(\omega_i) = -E(\omega_i) = \max|E(\omega)|,$$

其中 $\omega_0 < \omega_1 < \omega_2 < \cdots < \omega_{M+1}$ 且 $\omega_i \in F$.

零输入极限环特性(limit cycle property to a zero input)　无限单位脉冲响应(IIR)滤波器的一个基本特性. 对于无限长单位脉冲响应滤波器,其实现结构具有反馈环节. 假设其算术运算精度是无限的,在 $n > n_0$ 时刻,输入信号的幅度为零,则该滤波器的输出在 $n > n_0$ 时刻将逐渐衰减而趋向于零. 而同一滤波器,若以有限字长运算精度来实现,则该滤波器的输出在 $n > n_0$ 时刻可能衰减到某一非零的幅度范围,尔后呈现振荡特性. 它是 IIR 滤波器反馈环路中存在许多非线性量化的结果.

叠加原理(principle of superposition)　信号系统分析中的一个基本概念. 如果 $T[\cdot]$ 表示系统的输出与输入之间的变换关系,则对于任意两个输入 $x_1(n)$ 和 $x_2(n)$,以及任意标量 a 和 b,存在以下关系:

$$T[ax_1(n)+bx_2(n)]=aT[x_1(n)]+bT[x_2(n)],$$

则称该系统满足叠加原理. 叠加原理是判定线性系统的依据,同时为线性系统分析带来方便.

广义叠加原理(generalized principle of superposition)　叠加原理的推广. 如果用 $H[\cdot]$ 表示系统输入与输出之间的变换关系,符号"□"表示几个输入彼此组合的运算规则,符号"○"表示几个输出彼此组合的运算规则,符号":"表示输入与标量组合的运算规则,符号"⌉"表示输出与标量组合的运算规则,当系统满足关系

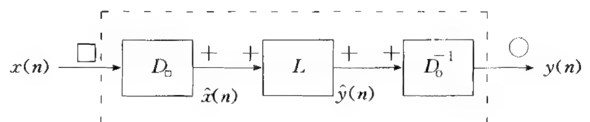
$$\begin{cases} H[x_1(n)\square x_2(n)] = H[x_1(n)]\bigcirc H[x_2(n)], \\ H[C : x_1(n)] = C \rceil H[x_1(n)], \end{cases}$$

则称该系统具有广义叠加性质,其运算服从广义叠加原理,所以是广义线性系统. 广义线性系统的变换属于抽象代数范畴,是线性矢量空间意义下的同态变换,所以通常把它称为同态系统. 同态系统能够把非线性问题转化为线性问题来处理,已经在图象处理、语音处理、地震预报等许多实际领域中得到重要应用.

同态系统(homomorphic system)　一种特定的数字信号处理系统. 根据广义叠加原理,如果把系统的输入与输出都解释为矢量空间中的矢量,把输入组合的运算规则"□"和输出组合的运算规则"○"对应于矢量加法,把输入与标量组合的运算规则":"和输出与标量组合的运算规则"⌉"对应于标量乘法,则系统的变换 $H[\cdot]$就是代数上的从输入矢量空间到输出矢量空间的线性变换. 由于这种变换是属于抽象代数范畴,是线性矢量空间意义下的同态变换,所以把具有这种输入与输出矢量空间之间的线性变换的系统称为同态系统. 又由于同态系统遵从广义叠加原理而且是可逆的,因此同态系统又称为广义线性系统.

在数字信号处理中它已经获得重要应用,特别是在处理乘法组合信号和卷积组合信号当中,显得非常有效. 通常,处理乘法组合信号的同态系统,称为乘法同态系统;处理卷积组合信号的同态系统,称为卷积同态系统.

同态信号处理(homomorphic signal processing)　一种特定的数字信号处理方法. 利用同态系统把一类输入信号变换为另一类信号输出的过程称



为同态信号处理. 例如,对输入为乘法组合的信号或者卷积组合的信号,通过同态系统就可以得到另一种组合的输出信号. 上图是表示同态系统的规范框

图. 图中，符号 $D_{\square}$ 表示运算为□的特征系统；$D_{\circ}^{-1}$ 表示运算为○的特征系统 $D_{\circ}$ 的逆系统；L 表示线性系统. 同态系统可以把非线性组合的信号，通过变换用线性系统对它进行处理.

倒谱(cepstrum)　一种描述信号序列的频谱结构. 把频谱的英文单词"spectrum"的前四个字母倒置，构成新的技术词汇——cepstrum(倒谱). 定义为序列功率谱取对数后的傅里叶反变换，即

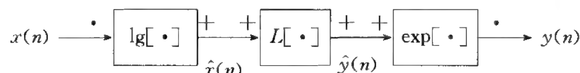
$$C(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln|X(e^{j\omega})|e^{j\omega n}\cdot d\omega.$$

可见倒谱具有时间因次，为了与通常频率函数的频谱相区别，有时称它为时谱.

乘法同态系统(multiplicative homomorphic system)　一种特定的数字信号处理系统. 它是专门处理乘法组合信号的同态处理系统. 乘法同态系统在语音处理和图象处理中都有具体应用. 乘法同态系统的数学原理如下：假设输入信号具有下面的组合形式：

$$x(n)=[x_1(n)]^{\alpha}\cdot[x_2(n)]^{\beta},$$

且对所有的 n，有 $x_1(n)>0, x_2(n)>0$，则输入和输出均为乘法运算的同态系统，其构成框图如图所示.



图中：

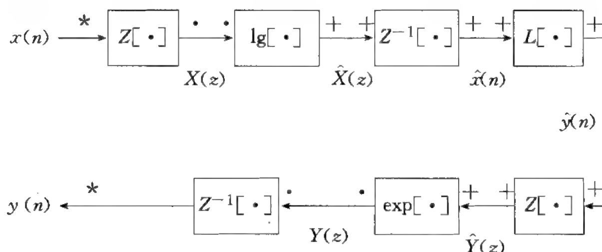
$$\begin{aligned}\hat{x}(n)&=\lg[x(n)]\\&=\lg[(x_1(n))^{\alpha}\cdot(x_2(n))^{\beta}]\\&=\alpha\lg[x_1(n)]+\beta\lg[x_2(n)]\\&=\hat{x}_1(n)+\hat{x}_2(n),\\\hat{y}(n)&=L[\hat{x}(n)]\\&=L[\hat{x}_1(n)+\hat{x}_2(n)]\\&=L[\hat{x}_1(n)]+L[\hat{x}_2(n)]\\&=\hat{y}_1(n)+\hat{y}_2(n),\\y(n)&=\exp[\hat{y}(n)]\\&=\exp[\hat{y}_1(n)+\hat{y}_2(n)]\\&=\exp[\hat{y}_1(n)]\cdot\exp[\hat{y}_2(n)]\\&=y_1(n)\cdot y_2(n).\end{aligned}$$

可见同态系统具有可分裂性，即信号通过该系统对于不同的成分提供可分别进行处理的可能性. 根据不同的应用目的，适当选择线性系统，便可得到不同的满意的处理结果.

卷积同态系统(homomorphic system for convolution)　一种特定的数字信号处理系统. 它是专门处理卷积组合信号的同态系统. 卷积同态系统在去混响和地质勘探中已经获得重要应用. 卷积同态系统的数学原理如下：假设输入信号具有如下组合形式：

$$x(n)=\sum_{k=-\infty}^{\infty}x_1(k)\cdot x_2(n-k)=x_1(n)*x_2(n),$$

则输入与输出均为卷积运算的同态系统，其构成方框图如图所示.



图中：

$$\begin{aligned}X(z)&=Z[x(n)]=Z[x_1(n)*x_2(n)]\\&=Z[x_1(n)]\cdot Z[x_2(n)]\\&=X_1(z)\cdot X_2(z),\\\hat{X}(z)&=\lg[X(z)]=\lg[X_1(z)\cdot X_2(z)]\\&=\lg[X_1(z)]+\lg[X_2(z)]\\&=\hat{X}_1(z)+\hat{X}_2(z),\\\hat{x}(n)&=Z^{-1}[\hat{X}(z)]=Z^{-1}[\hat{X}_1(z)+\hat{X}_2(z)]\\&=Z^{-1}[\hat{X}_1(z)]+Z^{-1}[\hat{X}_2(z)]\\&=\hat{x}_1(n)+\hat{x}_2(n),\\\hat{y}(n)&=L[\hat{x}(n)]=L[\hat{x}_1(n)+\hat{x}_2(n)]\\&=L[\hat{x}_1(n)]+L[\hat{x}_2(n)]\\&=\hat{y}_1(n)+\hat{y}_2(n),\\\hat{Y}(z)&=Z[\hat{y}(n)]=Z[\hat{y}_1(n)+\hat{y}_2(n)]\\&=Z[\hat{y}_1(n)]+Z[\hat{y}_2(n)]\\&=\hat{Y}_1(z)+\hat{Y}_2(z),\\Y(z)&=\exp[\hat{Y}(z)]=\exp[\hat{Y}_1(z)+\hat{Y}_2(z)]\\&=\exp[\hat{Y}_1(z)]\cdot\exp[\hat{Y}_2(z)]\\&=Y_1(z)\cdot Y_2(z),\\y(n)&=Z^{-1}[Y(z)]\\&=Z^{-1}[Y_1(z)\cdot Y_2(z)]\\&=Z^{-1}[Y_1(z)]*Z[Y_2(z)]\\&=y_1(n)*y_2(n).\end{aligned}$$

根据不同的应用目的，适当选择线性系统，便可得到不同的满意的处理结果. 值得注意的是，卷积同态系统在运算过程中为了避免出现相位多值性的问题，隐含着许多重要的假设. 首先，复对数运算

$$\hat{X}(z)=\lg[X(z)]=\hat{X}_1(z)+\hat{X}_2(z)$$

必须是确实成立的 Z 变换. 其次，Z 反变换运算

$$\hat{x}(n)=Z^{-1}[X(z)]=\hat{x}_1(n)+\hat{x}_2(n)$$

必须具有惟一性，即必须选择 $\hat{X}(z)$ 的定义域. 对于 $x(n)$ 和 $\hat{x}(n)$ 都是稳定的实序列时，便不会出现问题. 如果 $x(n)$ 和 $\hat{x}(n)$ 是不稳定的，则需要适当修正，使得 $X(z)$ 和 $\hat{X}(z)$ 的定义域都包含单位圆. 这样才能保证复对数的惟一性和解析性. 如果 $X(z)$ 和 $\hat{X}(z)=\lg X(z)$ 在单位圆上有定义，则 $X(e^{j\omega})$，

$X_1(e^{j\omega}),X_2(e^{j\omega})$必为 ω 的连续函数. 因为

$$X(e^{j\omega})=|X(e^{j\omega})|\cdot e^{j\arg X(e^{j\omega})},$$

所以 $\hat{X}(e^{j\omega})=\lg|X(e^{j\omega})|+j\arg X(e^{j\omega})$,

其中

$$\begin{cases}\lg|X(e^{j\omega})|=\lg|X_1(e^{j\omega})|+\lg|X_2(e^{j\omega})|,\\ \arg[X(e^{j\omega})]=\arg[X_1(e^{j\omega})]+\arg[X_2(e^{j\omega})].\end{cases}$$

这时,应选择$[-\pi,\pi]$范围内总的相位主值曲线. 而总的相位主值不一定等于各分量相位主值之和,因此,最后的结果得到的 $\arg[X(e^{j\omega})]$常常是$[-\pi,\pi]$范围内的逐段连续的相位曲线. 严格地解释上述问题,可以通过黎曼曲面理论.

复倒谱(complex cepstrum) 倒谱概念的推广. 如果信号是复数序列,则需采用复对数和复傅里叶变换,这样信号复对数功率谱的复傅里叶反变换便称为复倒谱. 复倒谱具有相位信息. 如果信号是实序列,其复倒谱也是实数,这时的复倒谱就是倒谱了.

频率取样设计(design by frequency sampling) 有限单位脉冲响应(FIR)滤波器的一种设计方法. 这种设计方法是根据频域取样定理,对所要求的频率特性进行取样,从样点中恢复原来的频率特性,达到设计数字滤波器的目的. 设所需的频率特性为 $H(e^{j\omega})$,若在一周期内对它作 N 点均匀取样,则得

$$H(e^{j\omega})|_{\omega=\frac{2\pi}{N}k}=H(k)\quad(k=0,1,\cdots,N-1).$$

通过离散傅里叶变换,最后求得 FIR 数字滤波器的系统函数为

$$H(Z)=\frac{1-Z^{-N}}{N}\sum_{k=0}^{N-1}\frac{H(k)}{1-e^{j\left(\frac{2\pi}{N}\right)k}X^{-1}}.$$

令 $Z=e^{j\omega}$,代入上式,得到频率响应为

$$H(e^{j\omega})=\frac{e^{j\omega(N-1)/2}}{N}\sum_{k=0}^{N-1}H(k)(-1)^k e^{-j\pi k/N}\cdot\frac{\left[N\left(\omega-\frac{2\pi}{N}k\right)2\right]}{\sin\left[N\left(\omega-\frac{2\pi}{N}k\right)2\right]}.$$

功率谱估计(power spectrum estimation) 平稳随机信号在频域各频率成分的功率分布情况估计. 功率谱估计是数字信号处理中的一个重要应用方面,在设计最佳线性滤波器、测量噪声频谱、检测淹没在宽带噪声中的窄带信号,以及用噪声激励法估计线性系统参数等,都要求通过功率谱估计,提高分辨率. 功率谱估计方法很多,大体可分为经典法和近代法两大类.

周期图法(periodogram method) 功率谱估计的一种经典方法. 由于平稳随机信号的自相关与功率谱是一对傅里叶变换,因此对有限长序列的功率谱可推得

$$I_N(\omega)=\frac{1}{N}|X(e^{j\omega})|^2.$$

将数字频率 ω 离散化,得到

$$I_N(k)=\frac{1}{N}|X(k)|^2\quad(k=0,1,\cdots,N-1),$$

式中,

$$X(k)=\sum_{k=0}^{N-1}x(n)e^{-j\frac{2\pi}{N}kn}$$

是有限长序列 $x(n)$的离散傅里叶变换,是以 N 为周期的周期函数. 所以把这种通过求$X(k)$计算功率谱的方法称为周期图法. 周期图法是功率谱的有偏估计. 也可解释为自相关序列加窗处理之后的傅里叶变换.

数字信号(digital signal) 信号的一种形式. 幅度量化、时间离散化的信号,称为数字信号. 严格地说,离散信号经过量化编码以后,才称之为数字信号. 如果采用二进制编码,则称为二进制编码数字信号. 编码数字信号中的每一个码位称为码元信号. 在数字通信系统中,实际传输的就是码元信号. 但是,在通信领域,人们常常把这些码元也称为数字信号.

左边序列(left-sided sequence) 序列的一种形式. 如果序列 $x(n)$ 在某一离散时刻 $n=n_0$ 以后等于零,即满足 $x(n)=0,n>n_0$,则 $x(n)$ 称为左边序列,其中 n_0 为常数.

右边序列(right-sided sequence) 序列的一种形式. 如果序列 $x(n)$ 在某一离散时刻 $n=n_0$ 以前等于零,即满足 $x(n)=0,n<n_0$,则 $x(n)$ 称为右边序列,其中 n_0 为常数.

双边序列(two-sided sequence) 序列的一种形式. 如果序列 $x(n)$ 在 $n\to+\infty$ 和 $n\to-\infty$ 时,都有非零值存在,则 $x(n)$ 称为双边序列. 双边序列也可以看做是左边序列与右边序列之和.

周期序列(periodic sequence) 具有周期特性的序列. 如果对于所有的 n,都能满足

$$x(n)=x(n+N),$$

则 $x(n)$称为周期序列. 式中 N 为正整数,是该序列的周期长度.

偶序列(even sequence) 一种具有偶对称特性的序列. 如果实数序列 $x_e(n)$满足

$$x_e(n)=x_e(-n),$$

则 $x_e(n)$称为偶序列. 共轭对称的实序列均为偶序列.

奇序列(odd sequence) 一种具有奇对称特性的序列. 如果实数序列 $x_o(n)$满足

$$x_o(n)=-x_o(-n),$$

则 $x_o(n)$称为奇序列. 共轭反对称的实序列均为奇序列.

共轭对称序列(conjugate symmetric sequence)

一种具有共轭对称特性的复数序列. 如果序列$x(n)$满足

$$x(n) = x^*(-n),$$

则$x(n)$称为共轭对称序列,式中符号 * 表示复数共轭.

共轭反对称序列(conjugate antisymmetric sequence) 一种具有共轭反对称特性的复数序列. 如果序列$x(n)$满足

$$x(n) = -x^*(-n),$$

则$x(n)$称为共轭反对称序列,式中符号 * 表示复数共轭.

二维单位取样序列(two-dimensional unit-sample sequence) 一种基本序列形式. 定义为

$$\delta(m,n) = \begin{cases} 1 & (m=0,n=0), \\ 0 & (\text{其他}). \end{cases}$$

二维单位阶跃序列(two-dimensional unit-step sequence) 一种基本序列形式. 定义为

$$u(m,n) = \begin{cases} 1 & (m \geqslant 0,n \geqslant 0), \\ 0 & (\text{其他}). \end{cases}$$

二维指数序列(two-dimensional exponential sequence) 一种基本序列形式. 形式为$a^m \cdot b^n$的序列,称为二维指数序列,其中a,b为任意常数. 二维指数序列是可分序列.

二维正弦序列(two-dimensional sinusoidal sequence) 一种基本序列形式. 形式为

$$A\cos(\omega_0 m + \varphi) \cdot \cos(\omega_1 n + \theta)$$

的序列,称为二维正弦序列. 其中幅度A、数字角频率ω_0,ω_1、相角φ,θ等均为常数. 二维正弦序列是可分序列.

可分序列(separable sequence) 具有可分离特性的序列. 如果二维序列$x(m,n)$能够分离成两个一维独立变量相乘积的形式,即

$$x(m,n) = x_1(m) \cdot x_2(n),$$

其中$m = 0 \sim M-1, n = 0 \sim N-1$,则$x(m,n)$称为可分序列. 可分序列的二维离散傅里叶变换也是可分离的,即

$$X(k,l) = X_1(k) \cdot X_2(l),$$

式中:$X_1(k)$为$x_1(m)$的M点一维离散傅里叶变换(DFT); $X_2(k)$为$x_2(n)$的N点一维 DFT. 对于可分序列,完全可用一维数字信号处理技术来进行处理.

共轭对称函数(conjugate symmetric function) 具有共轭特性的一种频谱函数. 序列的傅里叶变换,通常称为序列的频谱函数. 如果频谱函数$X(e^{j\omega})$满足

$$X(e^{j\omega}) = X^*(e^{-j\omega}),$$

则$X(e^{j\omega})$称为共轭对称函数. 式中符号 * 表示复数共轭. 偶序列的傅里叶变换,均为共轭对称函数.

共轭反对称函数(conjugate antisymmetric function) 具有共轭反对称特性的一种频谱函数. 序列的傅里叶变换,通常称为序列的频谱函数. 如果频谱函数$X(e^{j\omega})$满足

$$X(e^{j\omega}) = -X^*(e^{-j\omega}),$$

则$X(e^{j\omega})$称为共轭反对称函数. 式中符号 * 表示复数共轭. 奇序列的傅里叶变换,均为共轭反对称函数.

频率变换(frequency transformation) 设计滤波器的一种方法. 把一个频带范围变换到另一个相应的频带范围,称之为频率变换. 利用频率变换可设计各种类型的频率选择滤波器. 其基本思想是:根据给定的待设计指标,先变换成相应的低通原型滤波器的设计指标;然后按照低通滤波器的设计方法,设计出一个低通原型波滤器;最后再通过频率变换,求得所要求的实际低通、高通、带通、或带阻滤波器. 频率变换有模拟频率变换与数字频率变换,可根据设计对象的要求来选取.

排序(order arranging) 序列排列的方法. 指快速傅里叶变换算法流图中,输入序列和输出序列的排列顺序. 在基 2 快速傅里叶变换 (FFT)的流图中,输入序列与输出序列的排序规律一般是:当输入序列为正序排列,则输出序列为倒序排列;当输入序列为倒序排列,则输出序列为正序排列. 如果 FFT 的流图是以m为基底,将输入数据和输出数据的排列顺序标号,都用自然m进制编码表示时,则称为正位序排列,简称正序排列. 如果将其排列顺序标号都用自然m进制编码表示,然后将编码码位前后倒置,对所得到的新的排列顺序标号,称之为倒位序排列,简称倒序排列. 例如$N=8$,排列顺序标号用自然二进制编码,则正序排列为

000 001 010 011 100 101 110 111,

对应于十进制,则正序排列为

0 1 2 3 4 5 6 7.

对应于二进制,则倒序排列为

000 100 010 110 001 101 011 111,

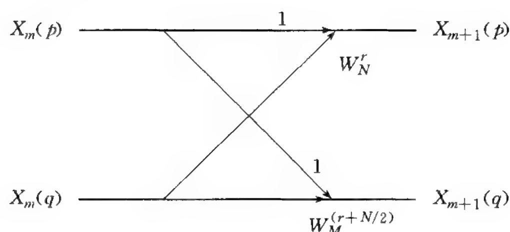
对应于十进制,则倒序排列为

0 4 2 6 1 5 3 7.

蝶形运算(butterfly computation) 一种实现快速傅里叶变换(FFT)运算的方法. 在基 2 快速傅里叶变换流图中,其结构都是由形似蝴蝶的运算单元所组成,所以称为蝶形运算. 如下图所示,其中

$$W_N^r = e^{j\frac{2\pi}{N}r}, \quad r = 0 ! \frac{N}{2} - 1.$$

蝶形运算(蝶算)的流图结构规范,编制计算程序时,便于实现同址运算,将大大节省计算机内存空间的占用量.



网络分析

网络分析(network analysis) 网络理论的一个分支学科.网络分析和网络综合是网络理论中的两个重要分支,两者互为补充.网络分析是指在给定了激励和网络结构的情况下,确定网络的响应.网络分析中,用器件和电路原理图来描述网络,并根据电路原理,从这个描述推导出网络在应用当中显示出什么样的性质.一般地,在一个网络分析问题中,几乎总是有一个惟一解,而不像在网络综合问题中,若有解,就是常有几个等效的解.网络分析可以用结点分析法、回路分析法以及状态空间法,而且既可以从时域出发也可以从频域出发进行网络分析.

网络综合(network synthesis) 网络理论的一个分支学科.它是网络分析的逆命题.网络综合是指在给定激励与需要的响应的情况下,求出物理上可以实现的网络函数,并进而求出能实现该响应的网络结构.网络综合包括两方面的问题,一是逼近问题,二是网络实现.前者是根据要求的响应求出物理上可实现的近似函数表示式,后者则是按此函数表示式实现其网络的结构.由于受到网络函数和网络元件在物理上可实现性的限制,所要求的响应并不总是能够实现的.如果能实现,由于满足一定要求的近似函数形式有多种,而同一函数又可用不同的综合方法得出不同的网络结构,故网络综合的结果不是惟一的.

由于是根据预给定的特性要求来实现其网络结构,所以,一般能保证所获得的网络结构具有较经济的元件数目,特别是能使设计结果尽量接近工作特性的要求,这正是网络综合能在现代科学领域中获得日益发展的主要原因.

网络(network) 各种物理实体连结成的一个特体结构.从广义上来讲,网络是对各支、点赋予了某种物理量的线图.从狭义上说,网络一般指电网络.就电网络而言,网络与电路二者在概念上没有严格的区别,因而在实际中这两个术语常常通用.

网络是由若干电气元件(如电阻器、电容器、电感器、变量器、晶体管、电源放大器等)相互连结构成,用以完成某一特定的功能,如滤波、延时、均衡等.完成上述功能的网络分别称为滤波网络、延时网络和均衡网络.

网络按照构成元件的类型和结构特征进行分类.含有能源的网络称为有源网络,不含能源的网络称为无源网络,由集总参数元件组成的网络称为集总参数网络,由分布参数组成的网络称为分布参数网络,只含有线性元件的网络称为线性网络,含有非线性元件的网络称为非线性网络,网络中只有两个端子的网络称为二端网络或单端对网络,有 N 个端子的网络称 N 端网络.

对于简单网络,或对复杂网络进行简化时,常运用网络定理或等效电路,对局部电路进行等效变换以简化分析.网络定理常用的有叠加定理、互易定理、戴文宁定理、诺顿定理和特勒根定理等,这些定理是分析网络的有力工具.

网络函数(network function) 一种特殊函数.指描述网络性能的函数.对于一个给定的线性无源网络,它的响应完全是由激励所确定的,两者间的因果关系代表着网络的工作特性,而这种因果特性常以网络函数来具体表示.对于多频率的信号,激励与响应都是频率的函数,并分别称为激励函数与响应函数,而将网络函数定义为响应函数与激励函数的函数比.采用复频率 s 为函数的自变量符号,可表示为

$$\text{网络函数 } H(s)=\frac{\text{响应函数 } R(s)}{\text{激励函数 } E(s)}.$$

网络函数根据研究的激励与响应是否在同一端对又可以分为两大类:第一类的激励与响应属于同一端对,这时的网络函数称为输入函数(或策动函数),如输入阻抗函数、输入导纳函数等;第二类的激励与响应属于不同端对,这时的网络函数称为转移函数(或传输函数),如转移导纳函数、转移阻抗函数、电压传输函数、电流传输函数等.

网络参数(network parameter) 能够独立地反映网络特性的参数.网络参数通常是指四端网络的参数,它包括 Z 参数、Y 参数、A 参数、特性参数等.

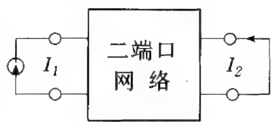
激励函数(exciting function) 指在网络的输入端加的激励信号.例如,电流源输出的电流信号、电压源输出的电压信号等,都可以作为网络的激励函数.

响应函数(response function) 在网络的输入端加上激励信号,则在网络的输出端便可获得相应的响应信号.该响应信号称为响应函数.响应函数可以是电压响应函数,可以是电流响应函数,也可是功率响应函数.

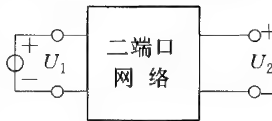
转移函数(transfer function) 描述网络性能的一种函数.对于一个给定的网络,如果输出端与输

入端分别是两个不同的端对，则输出端响应函数与输入端激励函数之间的比值，称为转移函数. 转移函数通常包括转移阻抗函数、转移导纳函数、转移电流比、转移电压比等.

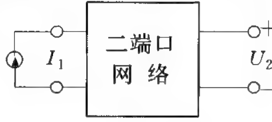
转移电流比(transfer current ratio) 网络输入端函数和输出端函数的一种关系. 如果网络的输出端与输入端分别是两个不同的端对，则输出端的电流响应函数与输入端的电流激励函数之比值，称为转移电流比. 如图，转移电流比为 I_2/I_1.



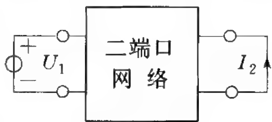
转移电压比(transfer voltage ratio) 网络输入端函数和输出端函数的一种关系. 如果网络的输出端与输入端分别是两个不同的端对，则输出端的电压响应函数与输入端的电压激励函数之比值，称为转移电压比. 如图，转移电压比为 U_2/U_1.



转移阻抗函数(transfer impedance function) 网络输入端函数与输出端函数的一种关系. 如果网络的输出端与输入端分别是两个不同的端对，则输出端的电压响应函数与输入端的电流激励函数之比值，称为转移阻抗函数. 如图，转移阻抗函数为 U_2/I_1.

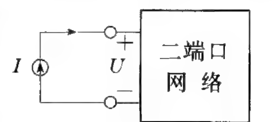


转移导纳函数(transfer admittance function) 网络输入端函数与输出端函数的一种关系. 如果网络的输出端与输入端分别是两个不同的端对，则输出端的电流响应函数与输入端的电压激励函数之比值，称为转移导纳函数. 如图，转移导纳函数为 I_2/U_1.



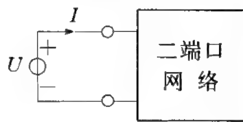
策动点函数(driving-point function) 即策动函数. 网络输入端函数与输出端函数的一种关系. 对于一个给定的网络，策动点是指网络的输入激励端对. 在网络中，任意指定的端对，都可以作为策动点. 在策动点处反映的网络函数，称为策动点函数. 也就是说，如果网络的输出端与输入端为同一端对，则输出端的响应函数与输入端的激励函数的比值，称为策动点函数.

策动点阻抗函数(driving-point impedance function) 描述网络性能的一种函数. 在网络的策动点处，输出电压响应与输入电流激励之间的比值，称为网络的策动点阻抗函数. 而网络类型可以是二端网络，也可以是四端网络；可以是无源网络，也可以是不含独立电源，只含有源元件的有源网络. 如图，给定一个二端口网络 N，它的策动点阻抗函数定义为



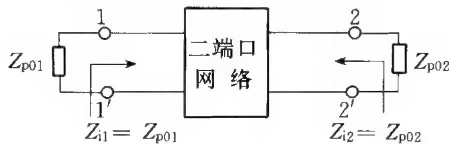
$$Z(s) = \frac{U}{I}.$$

策动点导纳函数(driving-point admittance function) 描述网络性能的一种函数. 在网络的策动点处，输出电流响应与输入电压激励之间的比值，称为网络的策动点导纳函数. 而网络类型，可以是二端网络，也可以是四端网络；可以是无源网络，也可以是不含独立电源，只含有源元件的有源网络. 如图，给定一个二端网络 N，它的策动点导纳函数定义为



$$Y(s) = \frac{I}{U}.$$

特性阻抗(characteristic impedance) 传输线的一种特性参数. 当传输线的终端负载阻抗等于某一特定阻抗时，负载将吸收掉传输到终端的全部能量，而无反射波产生，称这一特定阻抗为该传输线的特性阻抗. 这时从传输线上任一点投向传输线的输入阻抗都等于特性阻抗. 特性阻抗是由传输线本身的结构、参数和使用频率所决定的，与观察点的位置无关. 对于一个四端网络(或两端对网络)来说，其网络的特性阻抗也仅由网络本身元件值与具体结构所



决定，是网络所固有的参数，而与负载电阻、激励源内阻无关. 在电信号传输中，通常要研究传输线或网络的特性阻抗，以便分析信号传输中的阻抗匹配、信号相移与衰减等问题. 一般四端网络有两个特性阻抗 Z_{p01}, Z_{p02}，如在四端网络的输入端和输出端分别接上对应的特性阻抗，如图，则输入端和输出端都是阻抗匹配的，即 $Z_{i1}=Z_{p01}$, $Z_{i2}=Z_{p02}$.

开路阻抗(open-circuit impedance) 网络的一种特性参数. 对于"特性阻抗"中给定的四端网络，11′ 端的开路阻抗是指：当 22′ 端开路时，11′ 端的输入阻抗；22′ 端的开路阻抗是指：当 11′ 端开路时，22′ 端的输入阻抗(参见"特性阻抗").

短路阻抗(short-circuit impedance) 网络的一种特性参数. 对于"特性阻抗"中给定的四端网络，11′ 端的短路阻抗是指：当 22′ 端短路时，11′ 端的输入阻抗；22′ 端的短路阻抗是指：当 11′ 端短路时，22′ 端的输入阻抗(参见"特性阻抗").

网络参数矩阵(network parameter matrix) 描述网络参数的矩阵. 以矩阵形式表示的网络参数，

称为网络参数矩阵.例如，

Z 参数矩阵表示为 $\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$；

Y 参数矩阵表示为 $\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}$；

A 参数矩阵表示为 $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$.

参数矩阵中的每一个元素，都有固定的意义.

特性参数(characteristic parameter)　描述网络特性的参数.四端网络的特性参数，包括输入端对的特性阻抗、输出端对的特性阻抗和网络的传输常数.如图所示四端网络，设 11′端对的特性阻抗为 Z_{p01}，22′端对的特性阻抗为 Z_{p02}，网络的传输常数为 θ，则 Z_{p01}表示当 22′端对连结负载等于 Z_{p02}时，该网络在 11′端对的输入阻抗；Z_{p02}表示当 11′端对连结负载等于 Z_{p01}时，该网络在 22′端对的输入阻抗；θ 表示该网络的终端阻抗匹配连结的条件下，输入功率与输出功率之比，并取自然对数的二分之一，即

$$\theta = a + \mathrm{j}b = \frac{1}{2}\ln\left(\frac{I_1U_1}{I_2U_2}\right),$$

式中

$$a = \frac{1}{2}\ln\left|\frac{I_1U_1}{I_2U_2}\right|,$$

称为固有衰耗，单位为 dB，

$$b = \frac{1}{2}\arg(I_1U_1) - \frac{1}{2}\arg(I_2U_2),$$

称为固有相移，单位为 rad.

特性参数的计算公式为

$$Z_{p01} = \sqrt{Z_{10} \cdot Z_{1\infty}},\quad Z_{p02} = \sqrt{Z_{20} \cdot Z_{2\infty}},$$

$$\mathrm{th}\,\theta = \sqrt{\frac{Z_{10}}{Z_{1\infty}}} = \sqrt{\frac{Z_{20}}{Z_{2\infty}}},$$

其中 Z_{10}为 11′端对的短路阻抗；$Z_{1\infty}$为 11′端对的开路阻抗；Z_{20}为 22′端对的短路阻抗；$Z_{2\infty}$为 22′端对的开路阻抗.

归一化(normalization)　亦称标称化.网络分析术语.使网络标度统一，便于比较的一种方法.所谓归一化是在保证网络函数特性不变的条件下，改变网络函数的标度.归一化后，网络函数表示式简化，便于计算和网络设计.

频率归一化(frequency normalization)　一种归一化方法.选择某一频率 f_r 或角频率 $\omega_r(=2\pi f_r)$ 作为基准频率或基准角频率，将它除实际频率 f，就得到频率的归一化值，即

归一化频率：　$f' = \dfrac{f}{f_r}$，

归一化角频率：　$\Omega = \dfrac{\omega}{\omega_r} = \dfrac{f}{f_r}$.

归一化频率和归一化角频率数值相同，在网络分析和设计中多采用归一化角频率 Ω.

阻抗归一化(impedance normalization)　一种归一化方法.选取一个固定电阻 R_r 作为基准电阻，用基准电阻 R_r 去除实际阻抗 Z 就得到归一化阻抗，即归一化阻抗 $z = Z/Z_r$. 对 R,L,C 三类元件，阻抗归一化后分别为

$$\frac{R}{R_r} = r,$$

$$\frac{Z_l}{R_r} = \frac{sL}{R_r} = sl',$$

$$\frac{Z_c}{R_r} = \frac{1}{sCR_r} = \frac{1}{sc'},$$

式中 r,l',c' 分别为归一化电阻、归一化电感和归一化电容，它们与实际元件值关系为

$$r = \frac{R}{R_r},\ l' = \frac{L}{R_r},\ c' = CR_r.$$

所有 R 与 L 都用基准电阻 R_r 来除，而电容 C 乘以 R_r 表示各阻抗值均降低 R_p 倍.又可细分为

归一化电阻：　$r = \dfrac{R}{R_r}$，

归一化电感：　$l = \dfrac{L}{L_r}$，

归一化电容：　$c = \dfrac{C}{C_r}$，

式中 L_r 称基准电感，C_r 称基准电容.归一化值是没有量纲的，故 r,l',c' 又分别称为电阻系数、电感系数和电容系数.

复频率(complex frequency)　用复数来表示信号的频率和相位的一种方法.复频率 s 表示为 $s = \sigma + \mathrm{j}\omega$，其中复频率 s 的实部 σ 是用来描述信号中角度的增长或衰减的，用 dB/s 来测度，它的虚部 ω 通常称为实频率，用来描述信号的振荡，用 rad/s 来测度.复频率 s 使得人们在进行网络分析与综合的时候，可以利用复变函数这个数学工具，使分析与运算大大简化.

复平面(complex planar)　亦称 s 平面.网络综合理论中的平面.是用来标记复频率 s 的复数平面.横坐标为 σ，纵坐标为 $\mathrm{j}\omega$，$\sigma<0$ 的点在左半平面，$\sigma>0$的点在右半平面，$s=\sigma$ 的点在实轴 σ 上，$s=\pm\mathrm{j}\omega$的点在虚轴 $\mathrm{j}\omega$ 上，各种频率的指数信号都可以表示在 s 平面上，用复平面上各个复频率 s 点的位置来反映各种信号的形状，既方便，又明了，是网络综合理论中常用的图示方法.

网络函数的零点(zero-point of network function)　描述网络函数性能的一种参数.网络函数可以表示为两个 s 的实系数多项式之比.网络函数的

一般形式为

$$H(s)=\frac{A_ms^m+A_{m-1}s^{m-1}+\cdots+A_1s+A_0}{B_ns^n+B_{n-1}s^{n-1}+\cdots+B_1s+B_0},$$

式中分子多项式的根称为网络函数的零点，分母多项式的根称为网络函数的极点. 如果把无限大频率处的零点和极点也计算在内，零点和极点的数目应该相等.

网络函数的极点(pole-point of network function)　见“网络函数的零点”.

传输零点(zero-point of transmission)　网络函数的一个零点. 其物理意义是在这个频率的输出为零，即 $I_2=0$ 或 $U_2=0$. 对于 T 形网络，当任一串臂阻抗为无穷大时，信号传输路径中断，当任一并臂导纳为无穷大时，信号传输路径被短路，在这两种情况下，响应信号都可能变为零，即可以形成传输零点.

赫尔维茨多项式(Hurwitz polynomial)　简称赫氏多项式. 一种特殊多项式. 指只有左半平面零点的实系数多项式，它的零点只能位于左半平面和 $j\omega$ 轴上，且在 $j\omega$ 轴上的零点只能是单阶. 零点全部位于左半平面的赫氏多项式，称为严格赫氏多项式.

正实函数(positive real function)　一种实函数. 无源单端口网络的策动点函数必为正实函数才可以物理实现. 正实函数定义为：如果一个函数 $H(s)$ 是正实函数，则该函数满足下述条件：

1. 当 s 是实数时，$H(s)$ 是实数.

2. 当 s 的实部等于或大于零时，$H(s)$ 的实部等于或大于零，即 $\mathrm{Re}\,s\geqslant 0$ 时，$\mathrm{Re}[H(s)]\geqslant 0$.

戴文宁定理(Thevenin theorem)　亦称等效电源定理. 网络分析理论的基本定理之一. 即对于任何一个有源线性单端口网络对外电路而言，它可以用一个电动势为 E 的理想电压源和内阻 R_0 串联的电源来等效代替(图 1、图 2).

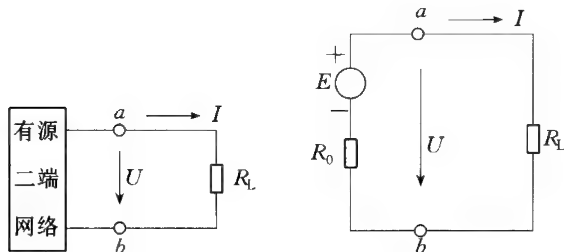


图1

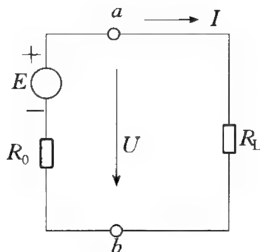


图2

图的等效电源的电动势 E 就是有源二端网络的开路电压 U_{ab}，即将负载断开后 a，b 两端之间的电压. 等效电源的内阻 R_0 等于有源二端网络中所有电源均除去(将各个理想电压源短路)，即电动势为零，将各个理想电流源开路(即电流为零)后所得到的无源网络 a，b 两端之间的等效电阻.

等效电路(图 2)是一个最简单的电路，其中电流可以由式

$$I=\frac{E}{R_0+R_L}$$

计算. 等效电源的电动势和内阻可通过实验或计算得到.

叠加定理(superposition theorem)　网络分析理论的基本定理之一. 它为线性网络中多种信号源同时激励时研究响应对激励的关系提供了理论依据和方法. 叠加定理指出：在任何线性网络中，任一支路的响应(电压或电流)是每个独立电源(包括初始条件源)单独作用时，在该支路中产生的响应的代数和. 由此可知：

1. 若要求出若干个独立源在网络中引起的响应，可以先求出每个独立源的响应而让其他所有源都不起作用，再计算其总和. 这一论点描述了线性网络的可加性.

2. 若网络中所有的电源都乘以一个常数，则其响应也应乘以同一个常数. 这一点又论述了线性网络的均匀性.

因此，叠加定理是线性网络的可加性和均匀性相结合的产物，是线性网络的线性特性的反映.

诺顿定理(Norton theorem)　网络理论的基本定理之一. 诺顿定理指出：任何一个单端口有源二端线性网络对外电路而言，它都可以用一个电流为 I_s 的理想电流源和内阻 R_0 相并联的电源来等效代替(图 1、图 2).

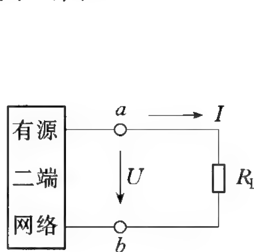


图1

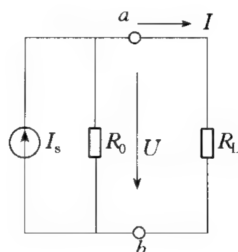


图2

等效电源的电流 I_s 就是有源二端网络的短路电流，即将 a，b 两端短接后其中的电流. 等效电源的内阻 R_0 等于有源二端网络中所有电源均除去(理想电压源短路、理想电流源开路)后所得到的无源网络 a，b 两端之间的等效电阻. 等效电路(图 2)中的电流可由式

$$I=\frac{R_0}{R_0+R_L}I_s$$

得到. 因此，一个有源二端网络既可用戴文宁定理化为等效电压源，也可由诺顿定理化为等效电流源. 两者对外电路来说是等效的，关系是

$$E=I_sR_0\quad\text{或}\quad I_s=E/R_0.$$

特勒根定理(Tellegen theorem)　网络理论的

基本定理之一.由于该定理的导出只依据基尔霍夫定律,因而与组成网络的元件性质无关,所以它的应用范围极为广泛,对于线性网络或非线性网络、时变和时不变网络都适用.特勒根定理有两种形式:

1.在一个具有 N 个结点和 B 条支路的集中参数网络中,令列向量 $[i_b]$ 和 $[u_b]$ 分别表示支路电流和支路电压,并将它们的参考方向取得一致,则

$$[u_b]^T[i_b]=0 \text{ 或 } \sum_{k=1}^{B}u_{bk}i_{bk}=0.$$

2.如果有两个网络 N 和 $\hat{N}$,它们是由不同的二端元件所构成,而它们的线图完全相同,用 $[i_b]$,$[u_b]$ 和 $[\hat{i}_b]$,$[\hat{u}_b]$ 分别表示这两个网络的支路电流和支路电压,则可写出

$$[u_b]^T[\hat{i}_b]=0 \quad \text{及} \quad [\hat{u}_b]^T[i_b]=0.$$

基尔霍夫定律(Kirchoff law) 亦称结点定律.电路理论的基本定律.是求解网络问题的基本工具,它包括第一定律与第二定律.基尔霍夫第一定律,它确定了电路中结点处电流间的关系,所以也称为电流定律.该定律指出:在电路中,任意时刻流入任意一个结点的电流必定等于流出该结点的电流;或描述为汇合于任一结点处的各电流的代数和等于零,即

$$\sum_{k=1}^{n}I_k=0.$$

通常规定流出结点的方向为正,流入的为负.基尔霍夫第一定律是"电荷守恒定律"的推广,是电流连续性的体现.从此定律获得了结点分析法.基尔霍夫第二定律,亦称回路定律.它确定了回路中电压间的关系,因此也称为电压定律.该定律指出:电路中任一回路内电压降等于电压升;或描述为:对任一闭合回路,各段电路上电压的代数和为零,即

$$\sum_{k=1}^{m}U_k=0.$$

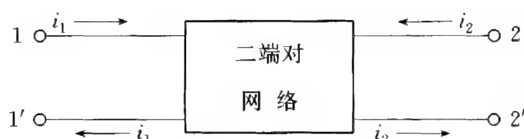
通常需要规定一个绕行回路的方向,凡电压的参考(或实际)方向与回路绕行方向一致者为正电压,否则为负电压.电源电压的参考方向为从电源的正极指向电源的负极.基尔霍夫第二定律是"能量守恒定律"的推广,是电压与路径无关性质的反映.从此定律出发,获得了回路分析法.

结点定律(node law) 见"基尔霍夫定律".

回路定律(loop law) 见"基尔霍夫定律".

互易定理(reciprocity theorem) 网络理论的基本定理之一.即对于一个线性时不变的无源网络,且初始条件也为零(即电容上初始电压和电感中初始电流为零),若在激励与响应位置互易时网络的结构不变,则 m 支路中的电动势 E_m 在 k 支路中所引起的电流 $I_k=E_m g_{km}$ 等于 k 支路中的电动势 E_k(与 E_m 数值相等)在 m 支路中所引起的电流 $I_m=E_k g_{mk}$,式中系数 g 具有电导量纲,称为转移电导.例如 g_{km} 是 k 支路和 m 支路的转移电导.在数值上,转移电导 g_{km} 等于在 m 支路的单位电动势作用下于 k 支路所产生的电流.

二端对网络(two-terminal pair network) 亦称四端网络.网络的一种类型.一个任意复杂的电路,如果它向外连结只有两对端子,而且一对是输入端子,另一对是输出端子,就称这样的网络为二端对网络或四端网络.用二端对网络的概念分析电路时,可以把这两对端子之间的电路概括在一个方框中,如图所示.



这个方框中的电路可能比较简单,也可能十分复杂.但根据基尔霍夫第一定律,在任何瞬时从某个端对的一个端子流入网络的电流,必然等于在同一瞬时从这一端对的另一个端子流出的电流.因此,二端对网络常被看做一个传输信息与能量的机构.电子线路中的各种放大器、变压器、RC 滤波网络、串联和并联谐振回路等,都是二端对网络的具体实例.

四端网络(four-terminal network) 即"二端对网络".

对偶网络(dual network) 网络的一种类型.如果有两个网络 N_1 和 N_2,它们的支路一一对应,其中一个网络的支路电压或电流的每一种约束关系,是另一个网络中相应支路把电压与电流互换之后的约束关系,则这两个网络互为对偶网络.例如,网络 N_1 中每一个支路电压与电流的约束关系满足关系式

$$f\left(v_1,i_1,t,\frac{\mathrm{d}}{\mathrm{d}t}\right)=0,$$

而网络 N_2 中,每一个支路电压与电流的约束关系满足关系式

$$f\left(i_2,v_2,t,\frac{\mathrm{d}}{\mathrm{d}t}\right)=0,$$

这时,网络 N_1 和 N_2 互为对偶网络.

等效网络(equivalent network) 网络的一种类型.即具有不同的网络结构和元件值,而网络特性相同的网络.它们对于外电路来说是等效的,这一特性称为网络等效.这样的网络称为等效网络.等效网络是可以相互取代的.

平面网络(planar network) 网络的一种类型.将网络结构画在平面上,总能使各条支路除了端点以外,不与其他支路相交,这种网络称为平面网络.实际电路,甚至非常复杂的电子线路,通常也都能用

平面网络表示. 而网络分析的许多定理和分析方法，都是针对平面网络的，甚至有些分析方法，比如网孔电流分析法，只能适用于平面网络.

非平面网络(non-planar network)　网络的一种类型. 将网络结构画在平面上，无论怎么画，也得不到平面网络，即网络中总会有某些支路，它除了端点以外，还与其他支路相交，这种网络称为非平面网络或立体交叉网络. 实际电路中，有些个别情况只能用非平面网络表示，比如传输线路中的交叉. 对于非平面网络的分析，有些电路分析方法是不适用的，比如网孔电流分析法. 这一点应当引起注意.

网络变换(network transformation)　不同类型网络之间的转换. 根据网络等效原理，网络的结构与元件值进行相应的变换，而保持网络特性不发生变化，这种变换称为网络变换. 例如，任何对称的无源模拟网络，都可以根据中剖定理，变换成格型网络. 任何数字网络，都可以根据转置定理，变换成另一种结构形式的数字网络. 这些都称为网络变换.

伯德图(Bode diagram)　一种表示网络函数频率特性的曲线. 网络函数的幅度频率特性和相位频率特性的曲线图，如果采用对数坐标，则称为波特图. 因为伯德(Bode, H. W.)最先使用这种方便的方法来研究网络函数的频率特性，故将这种图形称为伯德图(参见本卷《控制理论》同名条).

近似问题(approximation problem)　亦称逼近问题. 研究实现近似理想特性网络的一种问题. 近似问题是根据给定的理想特性及不同的误差准则要求，从而得到符合要求的网络函数，即找出一个函数，其衰减或相位特性符合规定的要求，并能用无源和有源网络来实现.

滤波器(filter)　网络的一种类型. 即只让某一频带的信号通过而阻止其他频带的信号通过的网络. 信号能通过的频带称为通带，信号不能通过的频带称为阻带. 根据通带和阻带在频率轴上位置的不同，滤波器可以分为低通滤波器、高通滤波器、带通滤波器和带阻滤波器.

勃脱瓦兹滤波器(Butter Worth filter)　亦称最平幅度滤波器. 具有某种特定特性的滤波器. 这种滤波器采用的误差准则是泰勒级数误差准则，所得到的幅度特性的模平方表示式为

$$|H(\mathrm{j}\omega)|^2 = \frac{1}{1 + B_n\omega^{2n}}.$$

这种滤波器的幅频特性具有以下特点：

1. $|H(\mathrm{j}\omega)|^2$ 值随 ω 值的增加单调下降.

2. 在 $\omega=0$ 处理想特性与实际特性误差最小.

3. 这种滤波器的传输零点都在 $\omega=\infty$ 处，是一种全极点滤波器.

切比雪夫型滤波器(Chebyshev filter)　具有某种特定特性的滤波器. 切比雪夫型滤波器是采用切比雪夫误差准则得到的，又称为 C 型滤波器. 其幅度特性的模平方表示式为

$$|H(\mathrm{j}\omega)|^2 = \frac{1}{1 + \varepsilon^2 C_n^2(\omega)},$$

式中 $C_n(\omega)$ 为切比雪夫多项式，ε 为小于 1 的实数. 这种滤波器的幅频特性具有以下特点：

1. $|H(\mathrm{j}\omega)|^2$ 在通带范围内随 ω 的增长呈现等波动变换，即对理想特性呈现等偏离变化.

2. $|H(\mathrm{j}\omega)|^2$ 在阻带范围内呈单调变化.

3. 传输零点全部位于 $\omega=\infty$ 处，是一种全极点滤波器.

考尔型滤波器(Cauer filter)　简称 CC 型滤波器. 具有某种特定特性的滤波器. 它是根据椭圆函数逼近准则得到的一种滤波器，故又称为椭圆函数滤波器. 它的幅频特性的模平方形式为

$$|H(\mathrm{j}\omega)|^2 = \frac{1}{1 + \dfrac{P(\omega^2)}{Q(\omega^2)}}.$$

其幅频特性特点如下：

1. 在通带和阻带范围内都是等波动变化.

2. 既有非零有限值的传输零点，又有 $\omega=\infty$ 处的传输零点.

灵敏度(sensitivity)　用来量度网络的元件变化所引起网络性能偏离其标准值程度的量. 灵敏度是定量地表示这种偏离的一种参数，是衡量电路优劣程度从而做出选择的最重要的参数之一.

模拟系统(analog system)　用以研究信息理论和方法的系统之一. 输入信号与输出信号都是模拟信号的系统. 由模拟器件构成. 对于具有稳定性和因果性的线性非移变模拟系统，可用其单位冲激响应 $h(t)$ 来完全表征，也可用其传输函数 $H(s)$ 来完全表征.

状态空间法(state-space method)　利用状态空间来研究系统特性的一种方法. 一般有因果关系的问题都可以用状态空间法来求解. 一个系统在每一瞬时的输出，不仅与该系统在该时刻的系统输入有关，而且与该系统在该时刻以前的状态有关. 一个系统在 t_0 时刻的状态，是一组描述该系统所必须具备的最少量数据，利用这组数据，连同系统的模型，以及在 $t \geqslant t_0$ 时的外施激励，就能够完全确定 t_0 以后任何时刻系统的特性. 所以，可以利用由状态变量、输入激励函数和系统元件模型的参数所组成的一组一阶微分方程——状态变量方程，来求出状态变量，以便最终求得系统的输出响应，这样的分析方法就称为状态空间法. 状态空间法尤其适用于分析多输入多输出系统、线性时变系统，甚至非线性系统，而且由于利用了线性代数这个有力工具，把冗繁的数

学式表达得非常简明，并且把状态变量微分方程纳入一种统一的标准形式，也特别便于用计算机来求解.

信号流图(signal flow graph)　一种网络结构图.是由连结结点的有向支路构成的一种网络结构图.它描述了线性方程组变量间的因果关系.对于数字网络，采用信号流图来进行分析与综合是非常方便的，它能够清楚地反映出数字网络从输入到输出的每一个结点处所参与的数字运算和运算结果情况.

流图转置(flow graph transposition)　网络结构的一种变换方法.信号流图中，将所有支路的方向倒置，并且将输入与输出对应交换，这种变换称为流图转置.流图转置以后，其输入与输出之间的传输比不发生变化.所以，利用流图转置，可以在输入与输出不发生变化的条件下得到一种新网络结构形式.这种新的网络结构，称为转置流图.

转置定理(transposition theorem)　描述信号流图传输函数之间关系的一个定理.该定理断言：互为转置的信号流图，其输出与输入之间的传输函数是相同的.转置定理为信号流图的转置结构提供了理论依据.

源结点(source node)　信号流图中的一类结点.信号流图中，表示注入流图的外部输入或信号源的结点称源结点.

阱结点(sink node)　信号流图中的一类结点.信号流图中，只有引入支路，没有引出支路的结点，称为阱结点.

神经元网络(network of neuron)　一种复杂网络系统.1943年，美国心理学家麦卡洛克(McCulloch，W.S.)和数学家皮特斯(Pitts，W.)提出的利用神经元网络对信息进行处理的数学模型，以及1949年希博(Hebb)提出的神经元之间连结强度变化的学习规则，即希博规则，开创了神经元网络的新局面.神经元网络是由大量类似于神经元的处理单元相互连结而成的非线性复杂网络系统，它是在现代神经科学研究成果的基础上提出的，试图通过模拟大脑神经元网络处理、记忆信息的方式，完成人脑那样的信息处理功能，是非线性的并行结构.神经元网络的运行方式有前馈式网络和反馈式网络.前馈式网络采用分层网络结构形式，实现从输入层结点的状态空间到输出层状态空间的非线性映射，它广泛采用于模式分类、特征抽取等方面；反馈式网络采用相互连结型网络结构形式，所在结点既可以输入，又是计算单元，同时向外输出，它用于优化计算和联想记忆.

神经元网络的学习方法有无导师的希博规则、有导师的δ规则、霍普菲尔德能量最小准则、误差反向传播的广义规则、玻耳兹曼规则等.

麦卡洛克-皮特斯模型(McCulloch-Pitts model)　简称MP模型.一种神经元网络模型.它是在1943年，由美国心理学家麦卡洛克(McCulloch，W.S.)和数学家皮特斯(Pitts，W.)等提出的利用神经元网络对信息进行处理的数学模型，从此人们开始了对神经元网络的研究.该模型的主要目的是完成对神经元状态的描述.通过对大脑的分析，人们发现，从信息处理的功能看，神经元由以下特点：

1. 多输入单输出.
2. 突触(传递神经冲动的地方)兼有兴奋和抑制两种性能.
3. 能时间加权和空间加权.
4. 可产生脉冲.
5. 脉冲进行传递.
6. 非线性(有阈值).

MP模型这样描述神经元状态：每个神经元的状态 $S_i(i=1,2,3,\cdots,n)$ 只能取0或1，分别代表抑制和兴奋.每个神经元的状态由MP方程决定：

$$S_i = f\Big(\sum_j w_{ij}S_j - \theta_j\Big)\quad (i=1,2,3,\cdots,n),$$

式中 w_{ij} 是神经元之间的连结程度，$w_{ij}(i\neq j)$ 是可调函数，由学习过程来调整；θ_j 是阈值；$f(x)$ 是阶梯函数.

演化计算(evolutionary computation)　一种模拟自然进化过程的算法.它包括三种算法：

1. 遗传算法.由美国霍兰德(Holland，J.)创立，后由琼(Jong，K.de)、格里芬斯塔特(Grefenstette，J.)、戈尔德伯革(Goldberg，D.)和戴维士(Davis，L.)等人进行了改进.
2. 演化规划.它最早由美国弗格尔(Fogel，L.J.)、奥文斯(Owens，A.J.)和沃尔什(Walsh，M.J.)提出，最近由弗格尔(Fogel，D.B.)进行了完善.
3. 演化策略.是由德国雷切恩伯格(Rechenberg，I.)和施沃弗(Schwefel，H.P.)建立的.

上述三种方法都是借鉴生物界自然选择和自然遗传机制的随机搜索算法，它们有两个共同的特点：群体搜索策略和群体中个体之间的信息交换.这两个特点的优越性表现在：首先，演化算法在搜索过程中不容易陷入局部最优，即使在所定义的适应度函数是不连续的、非规则的或有噪声的情况下，它们也能以很大的概率找到全局最优解；其次，由于它们固有的并行性，演化算法非常适合于巨量并行机；再者，演化算法采用自然进化机制来表现复杂的现象，能够快速可靠地解决非常困难的问题；此外，它们容易介入到已有的模型中并且具有可扩展性.

遗传算法(genetic algorithm)　简称GA.一种优化方法.是一种严格模拟生命进化机制新近发展

起来的搜索优化方法. 它由美国霍兰德(Holland, J.)创立, 后由琼(Jong, K. de)、格里芬斯塔特(Grefenstette, J.)、戈尔德伯革(Goldberg, D.)和戴维士(Davis, L.)等人进行了改进. 遗传算法模拟了自然选择和遗传中发生繁殖、交配和突变现象, 从任一初始种群出发, 通过随机选择、交叉和变异操作, 产生一群新的更适应环境的个体, 使群体进化到搜索空间中越来越好的区域. 这样一代一代不断繁殖, 最后收敛到一群最适应环境的个体上, 求得问题的最优解. 由于遗传算法利用了生物进化和遗传的思想, 所以它有着与传统方法不同的特征: 首先, 它的处理对象是问题参数的编码集, 而不是参数本身, 这样用传统方法很难解决的问题, 遗传算法都能解决, 因为它不受限制条件(如函数的连续性、导数存在、单极值)的约束; 其次, 遗传算法在搜索空间中同时在很多点进行求解, 这样就减少了收敛于局部最小的可能; 同时也增加了处理的并行性; 此外, 遗传算法利用随机规则而不是确定性规则来引导搜索; 而且, 遗传算法易于同别的技术结合使用, 容易介入到已有的模型中, 具有可扩展性.

反向传播网络(back propagation network) 简称 BP 网络. 一种神经元网络. 基于神经元网络进行机器学习(机器获取新知识和新技能, 并识别现有知识)的一种方法. BP 网络是一种前馈式神经元网络, 这种网络是 1985 年提出. 在这种网络中, 有两种信号在流动: 一是工作信号, 它是施加输入信号后向前传播, 直到在输出端产生实际输出的信号, 是输入和权值的函数; 二是误差信号, 网络实际输出与理论输出间的差值即为误差, 它由输出端开始, 逐层向后传播. 设在第 n 次迭代中输出端的第 j 个单元的输出为 $y_j(n)$, 则该单元的误差信号为

$$e_j(n) = d_j(n) - y_j(n);$$

定义单元 j 的平方误差为 $\frac{1}{2}e_j^2(n)$, 则输出端总的平方误差的瞬时值为

$$\xi(n) = \frac{1}{2}\sum e_j^2(n);$$

其均值为

$$\xi_{av} = \frac{1}{N}\sum_n \xi(n).$$

式中 ξ_{av} 为学习的目标函数, 学习的目的是使 ξ_{av} 达最小, ξ_{av} 是所有权值和阈值以及输入信号的函数. BP 网络是当前应用最广的一种网络, 比如它已在图象识别、边缘检测、模式记忆、异或问题、对称性判别和 T-C 匹配等方面得到应用. 但它也有缺点, 即训练时间长, 易陷入局部最小.

霍普菲尔德模型(Hopfield model) 一种反馈神经元网络. 它有离散型和连续型两种, 由霍普菲尔德(Hopfield, J. J.)分别在 1982 年和 1984 年提出. 这两种反馈式的神经元网络用于联想记忆和解决最优化问题, 使神经元网络有突破性进展. 一般在进行计算机仿真时采用离散型, 而在硬件实现时采用连续型. 离散型霍普菲尔德神经元网络是一个单层的神经元网络, 每个结点都相互连结. 该网络有个神经元 $S_i(i=1,2,3,\cdots,n)$, 它的状态变化是

$$S_i = \begin{cases} 1 & \left(\sum w_{ij}S_j - \theta_j > 0\right), \\ 0 & \left(\sum w_{ij}S_j - \theta_j \leqslant 0\right), \end{cases}$$

即

$$S_i(t+1) = f\left(\sum_{i\neq j} w_{ij}S_j(t) - \theta_j\right),$$

式中 $f(x)$ 是[0,1]阶梯函数, w_{ij} 是网络间的连结程度, θ_j 是阈值. 要求网络对称连结, 即 $w_{ij}=w_{ji}$, $w_{ii}=0$(自身不连结). 对于连续型的霍普菲尔德模型, 设该网络由 N 个神经元互连而成, 但神经元的输出不再是离散值 0 和 1, 而可以在一个区间内变化. $g(u)$ 为神经元作用函数, 通常取以下形式:

$$g(u) = \frac{1}{[1+\exp(-u/u_0)]}.$$

这两种霍普菲尔德模型都可以使能量函数(即勒奥普洛夫函数)随时间单调下降. 离散型的霍普菲尔德模型主要用于联想记忆, 连续型的霍普菲尔德模型成功解释了流动售货员路径问题(TSP 问题), 但它也有缺点, 要对称连结, 内存开销较大.

柯含农自组织模型(Kohonen self-organizing model) 具有学习功能的一种神经元网络模型. 在基于神经元网络的学习中, 有一种非监督式学习. 通过它可以提取一组数据中的重要特征或某种内在规律性(如分布特征, 或按某种特征聚类). 自组织过程即是一种非监督学习, 柯含农自组织模型是由芬兰学者柯含农(Kohonen, T.)于 1972 年提出的. 该模型能形成簇与簇之间的连续映射, 起向量量化器的作用. 该模型的提出基于以下三点:

1. 对生物神经系统的观察.

2. 局部相互作用(可以是随机的)可以导致整体的某种有序性.

3. 这种局部作用遵循以下原则:

1) 突触(传递神经冲动的地方)的连结强度趋向于自增强.

2) 由于资源有限, 使各单元之间发生竞争, 并导致有选择性的变化.

3) 单个神经元不起作用, 所以可以说各神经元之间存在协作.

4) 人脑由大量的神经元组成, 它们并非都起同样的作用.

5) 处于空间不同部位的区域分工不同, 各自对

不同模式敏感.

因此,柯含农认为,神经元网络中邻近的各单元通过相互作用,可以自适应地发展成检测不同性质信号的特殊检测器.他又进一步提出一种学习方式,使得输入信号可以映射到低维空间,且保持相同特征的输入信号在空间上对应邻近区域,称为自组织特征映射.对该模型的使用实例是指纹识别.

模拟退火(analog anneal) 简称 AA.寻找全局最优解的一种方法.在神经元网络的研究中有许多地方涉及全局最优问题,寻优的过程是一个搜索过程,模拟退火过程就是一个寻找全局最优(使目标函数全局最小)的过程,它是模仿固体物质的退火过程(高温物质降温时其内能随之下降,至某一低温时内能最小),这种模拟退火过程的寻优方法称为模拟退火算法.模拟退火算法的步骤如下:

1.随机给定初始状态,选择合适的退火策略(温度下降规律),给初始温度 T_0 以足够高的值.

2.令 $x'=x+\Delta x$(Δx 为很小的均匀分布的随机扰动),并计算 $\Delta E=E(x')-E(x)$.

3.$\Delta E<0$,则接受 x' 为新的状态,否则以概率 $P=\exp(-\Delta E/kT)$ 接受 x',其中 k 是玻耳兹曼常数.具体作法是产生 0 到 1 之间的随机数 a,如 $P>a$,则接受 x' 为新的状态,否则仍留在原状态.

4.重复步骤 2 和 3 直至系统达到平衡态.

5.按步骤 1 给定的退火策略,重复步骤 2 至 4,直至 $T=0$ 或到达某一预定的低温.

以上介绍的是模拟退火的主要步骤.为了加快收敛速度,有不少人提出了改进方法,如快速 AA、自适应 AA、甚快速 AA 等.

数据处理组合算法(group method of data handling) 简称 GMDH.亦称感应学习算法,或自组织算法.一种模拟大脑进行过程的算法.常用于复杂系统分析.这种算法产生于感应器理论,其基础是自组织原则,它在模式识别、数学建模及对随机过程的预测方面有重要的用途.这种基于感应方法的算法模拟了人们大脑中进行的过程,科学家们将这种算法看做是模式识别理论、控制论、信息论、系统科学及其他一些学科的综合,用于环境系统分析、经济系统分析、农业系统分析,及对时间序列的评价等诸多方面.

数据处理组合算法的基本思想是:

1.所研究问题的样本数据容量为 N.

2.将此 N 个数据分成两个集合 N_A 和 N_B,其中 N_A 是培训集,N_B 是测验集.

3.在输入和输出变量间建立一个参考函数.

4.在标准的规则集中选择合适的目标函数,如偏差最小、一致性原则等.

5.以参考函数为基础,将不同的局部方程分类.

6.用培训集 N_A 和参数估计的方法对各局部方程的权重进行估计.

7.用测验集 N_B,根据选定的目标函数来对这些局部方程进行量度.

8.选择最好的测量方程作为最佳模型.

感应学习算法(inductive learning algorithm) 即"数据处理组合算法".

撰 稿 叶 明 叶亚斌 乐光新 刘 军 刘景堂 齐立心 孙 伟 李承恕 吴伟陵 吴德明 汪齐贤 张永辉 陈庆仁 孟秀林 赵荣黎 秦 涛 徐安士 徐志根 郭光灿 郭宇春 蒲祯德 戴无惧

审 阅 吴湘淇 沈世镒 张世演 张汉一 张家谋 赵尔源 胡正名 钟义信 袁保宗 郭光灿 章照止 程时端 蔡长年

画法几何与工程图学

画法几何(descriptive geometry)　几何学的一个分支.它是研究空间形体在平面上的图示法和在平面上解决空间几何问题的图解法的学科.空间形体,比如机器、房屋、桥梁等,它们的形体与构造都很复杂,必须按照一定的规律和方法,制出图样,才能表达清楚,使各处工程技术人员看了都能明白,并能据此制造施工.画法几何研究从客观物体中抽象出来的点、线、面、体,它们是客观形体的实在反映.画法几何摒弃了一切物体的自然属性,仅保留其几何形式,从而具有应用广泛性、高度概括性,为表达形体提供了理论基础.研究其他学科时,常需要解决一些空间问题,许多情况是可以应用简捷直观、基于画法几何理论的图解法.比如,作用在某一物体上空间各力的分析、机构空间位置的图解、船身车身型线的设计等.

空间想象力对于工程技术人员非常重要.画法几何除了应用在表达设计思想、认识形体构造外,对创造新产品、新技术以及培养创造性人才具有深刻意义.当设计一个新产品时,它们的形象首先产生在设计者的脑子里,然后需要运用更多的空间想象力.画法几何的学习和研究,能够有力地发展这方面的能力.

用图样表达人们的思想,早在远古时代就在人类实践中产生.画法几何里的一些知识和方法,已在各地各个技术领域产生和应用.而画法几何成为一门完备系统学科,则是由法国几何学家蒙日(Monge,G.)所创立.在当时法国革命时代,技术迅速发展的形势下,他总结和发展了人类长期在本学科领域的劳动成果,于1795年写成本学科的第一本著作《画法几何学》,自此画法几何形成了数学应用几何的一个分支.

画法几何是以投影法为基础.投影来源于实际,经抽象而得出高度概括的规律和方法,从而建立了表达和认识形体的理论.画法几何研究的点、线、面,是组成形体几何轮廓的基本元素.物体形状变化万千,而组成其几何轮廓的不外是点、线、面.对点、线、面的研究与分析,从而可以得到高度概括的表达形体的理论.投影可以是在平面上,也可以在柱面及其他曲面上.投影法一般有正投影法、斜投影法、中心投影法、轴测投影法及圆弧投影法等.

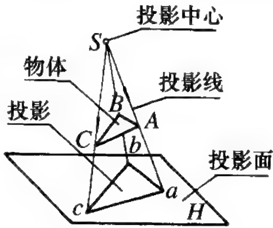
用射影几何原理来阐述画法几何内容,是进一步的研究,由此可逐步扩展到应用各种几何变换和其他特殊方法来处理画法几何问题.目前,应用计算机解决画法几何问题的研究,是一个新的发展趋势.

在本分支中,只介绍画法几何的基本条目.

投影法与投影变换

投影法(projection method)　在平面(图纸)上表达空间物体形象的方法,是画法几何学的基础.在日常生活中可以见到一些投影现象.物体在光源(日光或灯光)的照射下,就会出现物体的影子,投影方法就是从这些自然现象抽象出来的.画法几何学就是靠这种投影法,确定空间物体在平面上(图纸上)的图象."中心投影法"条目中的图是三角板投影的例子.

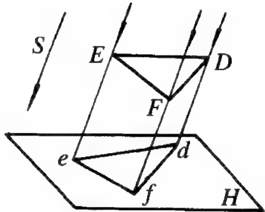
中心投影法(method of central projection)　一种基本、常用的投影法.在投影法中,所有投影线都通过一点——投影中心,称为中心投影法.如图,设点 S 表示光源,称为投影中心,光线称为投影线,即图中的 SA,SB,SC.地面或桌面称为投影面,用字母 H 表示.SA,SB,SC 的延长线与 H 面交于 a,b,c 三点.这三点就是空间点 A,B,C 在 H 面上的中心投影.



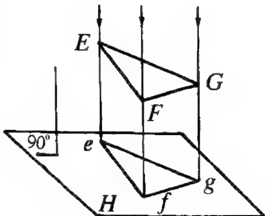
投影中心(projection center)　见"中心投影法".

投影线(projection line)　见"中心投影法".

平行投影法(method of parallel projection)　一种基本、常用的投影法.若将中心投影法中的投影中心移至无穷远,则各投影线相互平行,此时,物体在投影面上同样也得到投影,这种投影法称为平行投影法.如图所示,图中 S 表示投影方向.



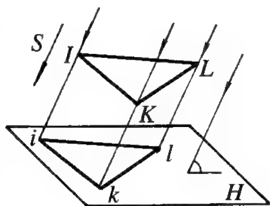
直角投影(orthographic projection)　亦称正投影.平行投影的一种.若投影方向垂直于投影面,则称这种投影为直角投影.如图所示.



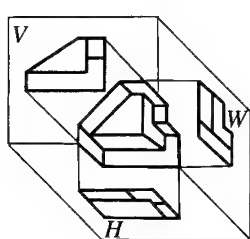
正投影(orthographic projection)　即"直角投

影".

斜投影(oblique projection)　平行投影的一种. 若投影方向 S 倾斜于投影面,则称这种投影为斜投影. 如图所示. 斜投影应用较少,但有时它绘出的图样具有很好的效果. 如斜轴测投影,对某些物体既能显示主要面的形象、有立体感,又可便利地绘出轴测图.



正投影法(method of orthographic projection)　一种多面投影. 用两个或两个以上相互垂直的投影



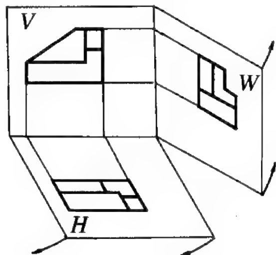
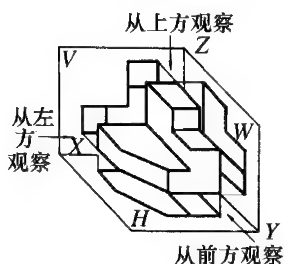


图 1　　　　图 2

面,在每个投影面上,分别做出空间几何体的直角投影,然后将互相垂直的各投影面按一定的方法展平在同一平面上,由这些投影便能完全确定空间几何体的形状和位置,这种投影法称为正投影法. 正投影法是表达复杂空间关系的一种准确方法,在机械制造行业和其他工程部门中被广泛采用. 其缺点是直观性较差. 图示是几何体的正投影.

第一角投影法(first-angle projection method)　一种正投影法. 在如图 1 所示的互相垂直的 V,H,W 三投影面构成的三投影面体系中,求几何体在该



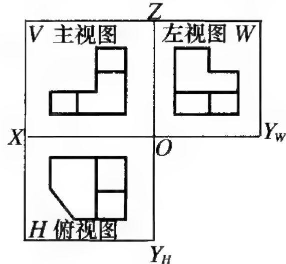
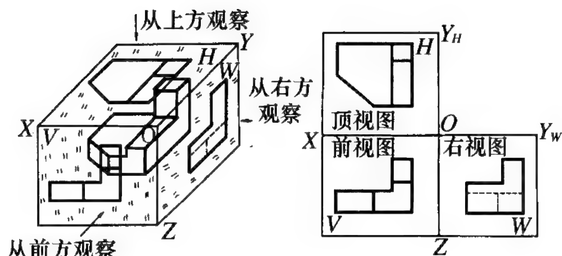


图 1　　　　图 2

体系中的投影称为第一角投影法. 第一角投影法的投影顺序是:观察者→几何体→投影面. 各视图的配置如图 2 所示. 在 V 面上的投影称为主视图,在 H 面上的投影称为俯视图,在 W 面上的投影称为左视图.

中国采用第一角投影法.

第三角投影法(third-angle projection method)



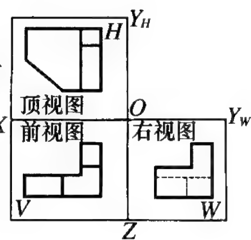


图 1　　　　图 2

一种正投影法. 在图 1 所示的互相垂直的 V,H,W 三投影面体系中,求几何体在该体系中的投影称为第三角投影法. 第三角投影法的投影顺序是:观察者→投影面→几何体. 各视图的配置如图 2 所示. 视图的名称不同于第一角投影法,在 V 面上的投影称为前视图,在 H 面上的投影称为顶视图,在 W 面上的投影称为右视图.

采用第三角投影法的国家有美国、加拿大等.

两种投影法的标志(sign of two projection methods)　国际标准的一项规定. 在国际标准(ISO)中规定,可以采用第一角投影法,也可以采用

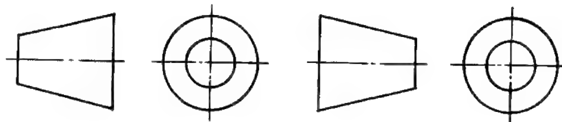
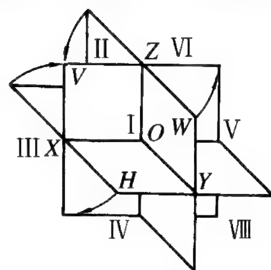


图 1　第一角画法　　　　图 2　第三角画法

第三角投影法. 为了区别两种画法,应在标题栏内用规定标志符号表示. 如图所示.

投影面(projection plane)　投影法的基本概念之一. 在投影法中,承受投影的平面或其他表面称为投影面. 通常设三个互相垂直的平面为投影面,其中 V 面称为正立投影面,H 面称为水平投影面,W 面称为侧立投影面. 三个互相垂直的投影面,将空间分为八个部分,每一部分称为一个分角. W 面左侧空间为第Ⅰ、Ⅱ、Ⅲ、Ⅳ分角,W 面右侧空间为Ⅴ、Ⅵ、Ⅶ、Ⅷ分角,其顺序如图所示.



正立投影面(frontal projection plane)　见"投影面".

水平投影面(horizontal projection plane)　见"投影面".

侧立投影面(profile projection plane)　见"投影面".

分角(angle of division)　见"投影面".

投影轴(projection axis)　投影法的基本概念

之一. 投影面的交线称为投影轴. 例如, V 和 H 面的交线为 OX 轴; H 和 W 面的交线为 OY 轴; V 和 W 面的交线为 OZ 轴. 三轴相交于一点 O, 称为原点.

图 1　　图 2

其中 X 轴的左方、Y 轴的前方、Z 轴的上方为正值; X 轴的右方, Y 轴的后方, Z 轴的下方为负值. 如图所示.

原点(origin)　见"投影轴".

投影连系线(projecting relative line)　投影法的基本概念之一. 几何元素(点、线、面)或几何体各投影之间的连线称为投影连系线. 如图所示.

点的投影(projection of point)　一种最基本的投影. 指点的直角投影. 在三投影面体系中(如图 1), 由空间点 B 分别向三个投影面作垂线, 垂线与各投影面的交点, 称为点的投影. 在 V 面上的投影称为正面投影, 以 b' 表示; 在 H 面上的投影称为水平投影, 以 b 表示; 在 W 面上的投影称为侧面投影, 以 b''表示. 然后, 将投影面进行旋转, V 面不动, H, W 面按箭头方向旋转 90°, 即将三个投影面展成一个平面, 从而得到点的三个投影的正投影图, 如图 2 所示.

图 1　　图 2

点的投影有如下规律:

1. 点的正面投影 b' 与水平投影 b 的连线垂直于 OX 轴;

2. 点的正面投影 b' 与侧面投影 b''的连线垂直于 OZ 轴;

3. $bb_x=b''b_z$, 即点的水平投影 b 到 X 轴的距离等于点的侧面投影 b''到 Z 轴的距离.

点的正面投影(frontal projection of point)　见"点的投影".

点的水平投影(horizontal projection of point)　见"点的投影".

点的侧面投影(profile projection of point)　见"点的投影".

点的投影规律(rule of point projection)　见"点的投影".

点的坐标(coordinate of point)　空间点位置的表示. 以三个投影面作为坐标面, 三个投影轴作为坐标轴, 则空间点 B 到投影面的距离, 称为点的坐标, 点到 W 面的距离为 X 坐标, 到 V 面的距离为 Y 坐标, 到 H 面的距离为 Z 坐标. 如图所示. 用三个坐标确定点 B, 可标记为 $B(X,Y,Z)$. 从图中可以看出: $Bb'b_xb$, $Bb'b_zb''$, Bbb_yb''均为矩形, 故可得

$$Bb'' = b'b_z = bb_y = b_xO = x,$$
$$Bb' = bb_x = b''b_z = b_yO = y,$$
$$Bb = b'b_x = b''b_y = b_zO = z.$$

点的每个投影反映两个坐标, 即正面投影 b' 反映 X, Z 坐标; 水平投影 b 反映 X, Y 坐标; 侧面投影 b''反映 Y, Z 坐标.

重影点(projection-coincidence point)　投影法的基本概念之一. 当两点位于同一投影线上时, 则它们对某一投影面上的投影重合, 此两点称为对该投影面的重影点, 如图所示.

图 1　　图 2

图中 E, D 两点在 H 面上的投影 e, d 重合为一点, 故为重影点. 其中 E 点的 Z 坐标大于 D 点的 Z 坐标, 故 E 点为可见, D 点为不可见, 其水平投影 d 加括弧, 表示不可见. 同理, C, A 两点对 V 面为重影点, c', a' 重合为一点, C 点的 Y 坐标大于 A 点的 Y 坐标, 故 C 点的正面投影 c' 为可见, A 点的正面投影 a' 为不可见.

投影面平行线(parallel line of projection plane)　一类具有特殊位置的直线. 指平行某一投影面, 而与另外两个投影面倾斜的直线.

正平线(frontal line)　一种投影面平行线. 平

行于 V 面而与 H,W 面倾斜的直线称为正平线. 如图所示. 其投影特性为:

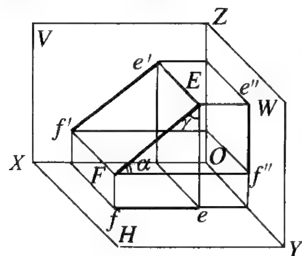


图 1

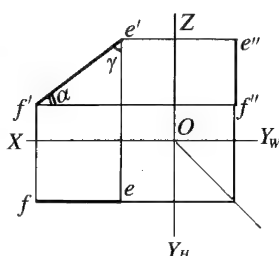


图 2

1. 正平线的正面投影反映该直线的实长,即 $e'f'=EF$;

2. 正平线的正面投影 $e'f'$ 与 OX 轴的夹角反映该直线对 H 面的倾角 α,与 OZ 轴的夹角反映该直线对 W 面的倾角 γ;

3. 正平线的水平投影平行于 OX 轴,即 $ef \parallel OX$,正平线的侧面投影平行于 OZ 轴,即 $e''f'' \parallel OZ$.

水平线(horizontal line) 一种投影面平行线. 平行于 H 面而与 V,W 面倾斜的直线称为水平线. 如图所示. 其投影特性为:

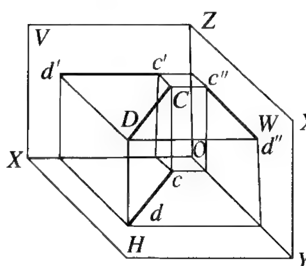


图 1

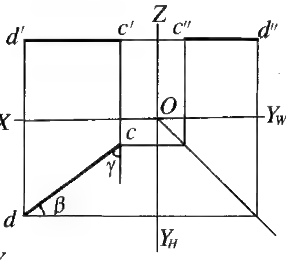


图 2

1. 水平线的水平投影反映该直线的实长,即 $cd=CD$;

2. 水平线的水平投影 cd 与 OX 轴的夹角反映该直线对 V 面的倾角 β,与 OY 轴的夹角反映该直线对 W 面的倾角 γ;

3. 水平线的正面投影平行于 OX 轴,即 $c'd' \parallel OX$,水平线的侧面投影平行于 OY_W 轴,即 $c''d'' \parallel OY_W$.

侧平线(profile line) 一种投影面平行线. 平行于 W 面而与 V,H 面倾斜的直线称为侧平线. 如图所示. 其投影特性为:

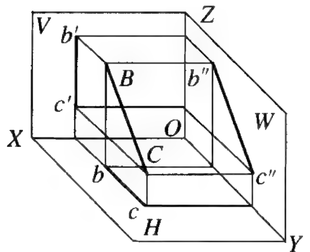


图 1

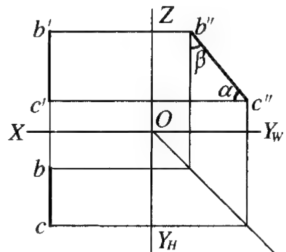


图 2

1. 侧平线的侧面投影反映该直线的实长,即 $b''c''=BC$.

2. 侧平线的侧面投影 $b''c''$ 与 OY_W 轴的夹角反映该直线对 H 面的倾角 α,与 OZ 轴的夹角反映该直线对 V 面的倾角 β.

3. 侧平线的正面投影平行于 OZ 轴,即 $b'c' \parallel OZ$,侧平线的水平投影平行于 OY_H 轴,即 $bc \parallel OY_H$.

投影面垂直线(perpendicular line of projection plane) 一类具有特殊位置的直线. 指垂直某一投影面,而与另外两个投影面平行的直线.

正垂线(line perpendicular to vertical plane) 一种投影面垂直线. 垂直于 V 面的直线称为正垂线. 如图所示. 其投影特性为:

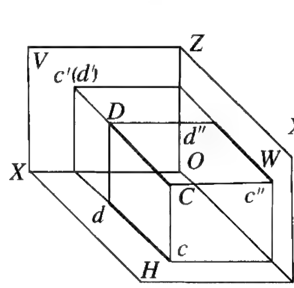


图 1

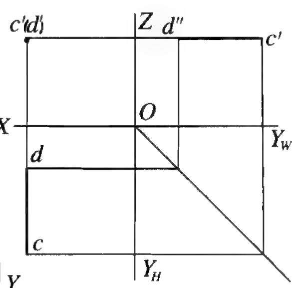


图 2

1. 正垂线的正面投影积聚为一点,即 c',d' 重合.

2. 正垂线的水平投影 cd 和侧面投影 $c''d''$ 反映该直线实长,即 $cd=c''d''=CD$.

3. 正垂线的水平投影垂直于 OX 轴,即 $cd \perp OX$,正垂线的侧面投影垂直于 OZ 轴,即 $c''d'' \perp OZ$.

铅垂线(line perpendicular to horizontal plane) 一种投影面垂直线. 垂直于 H 面的直线称为铅垂线. 如图所示. 其投影特性为:

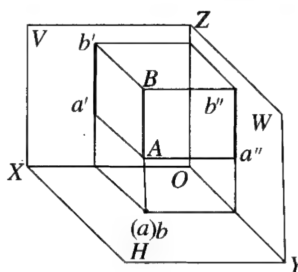


图 1

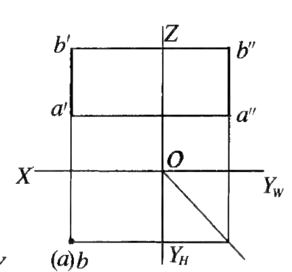


图 2

1. 铅垂线的水平投影积聚为一点,即 b,a 重合.

2. 铅垂线的正面投影 $b'a'$ 和侧面投影 $b''a''$ 反映该直线实长,即 $b'a'=b''a''=BA$.

3. 铅垂线的正面投影垂直于 OX 轴,即 $b'a' \perp OX$,铅垂线的侧面投影垂直于 OY_W 轴,即 $b''a'' \perp$

OY_W.

侧垂线(line perpendicular to profile plane) 一种投影面垂直线. 垂直于 W 面的直线称为侧垂线. 如图所示. 其投影特性为:

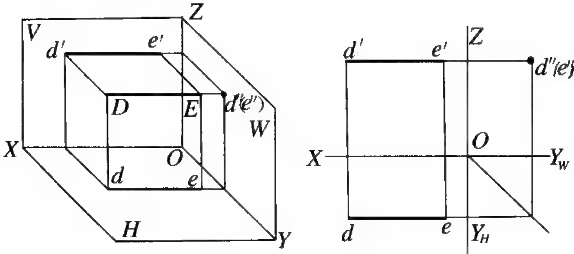


图 1　　　　图 2

1. 侧垂线的侧面投影积聚为一点,即 d'',e'' 重合;

2. 侧垂线的正面投影 $d'e'$ 和水平投影 de 反映该直线实长,即 $d'e'=de=DE$;

3. 侧垂线的正面投影垂直于 OZ 轴,即 $d'e'\perp OZ$,侧垂线的水平投影垂直于 OY_H 轴,即 $de\perp OY_H$.

一般位置直线(oblique line) 一种直线. 指对三个投影面无平行、垂直关系,而对三个投影面都倾斜的直线称为一般位置直线,如图所示. 直线 FA 与 H,V,W 三个投影面的夹角分别用 α,β,γ 表示,则

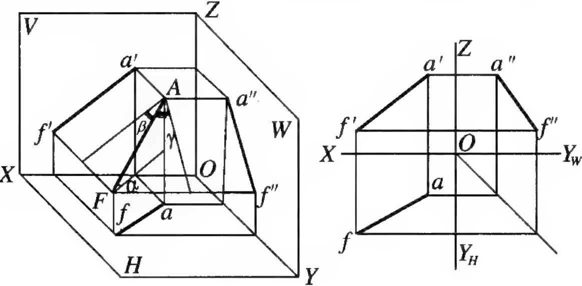


图 1　　　　图 2

$$fa = FA\cos\alpha,$$
$$f'a' = FA\cos\beta,$$
$$f''a'' = FA\cos\gamma.$$

由于直线 FA 倾斜于各投影面,故其各个投影长度均小于直线的实长. 其各个投影均与投影轴倾斜.

平行两直线的投影(projection of two parallel

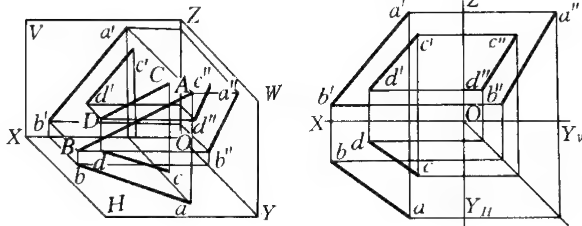


图 1　　　　图 2

straight lines) 一种基本、常见的投影. 指居于平行位置的两直线的投影. 空间两直线互相平行,则它们的投影也一定互相平行,即 $AB /\!/ CD$,则

$$a'b' /\!/ c'd',$$
$$ab /\!/ cd,$$
$$a''b'' /\!/ c''d''.$$

如图所示. 反之,若两直线的各个同面投影都互相平行,则该两直线平行.

相交两直线的投影(projection of two intersecting straight lines) 一种基本、常见的投影. 指居于相交位置的两直线的投影. 空间两直线相交,则它们的投影也一定相交,且交点符合点的投影规律.

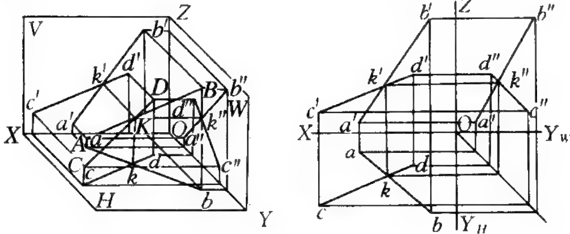


图 1　　　　图 2

如图所示. 反之,若两直线各个同面投影都相交,且交点符合点的投影规律,则该两直线在空间一定相交.

交叉两直线的投影(projection of two skew straight lines) 一种基本、常见的投影. 指居于交叉位置的两直线的投影. 空间两直线既不平行,也不相交,则称此两直线为交叉两直线. 如图所示.

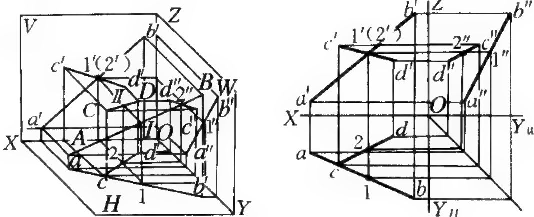


图 1　　　　图 2

交叉两直线的投影,可能有一个、两个或三个投影相交,但其交点不符合点的投影规律,即连结直线投影的两个交点,连线不与投影轴垂直. 图中 1,2 两点是对 V 面的重影点.

直线的迹点(trace point of straight line) 直线上的一种特殊点. 直线与投影面的交点称为直线的迹点. 如图所示. 直线 EF 与 V 面的交点称为正面迹点,以 N 表示;与 H 面的交点称为水平迹点,以 M 表示;与 W 面的交点称为侧面迹点,以 S 表示(图中未显示). 迹点既是直线上的点,又是投影面上的点,故其投影具有以下两个特点:

1. 因迹点是直线上的点,所以其投影必在直线的同面投影上(在同一个投影面上的投影,称为同面

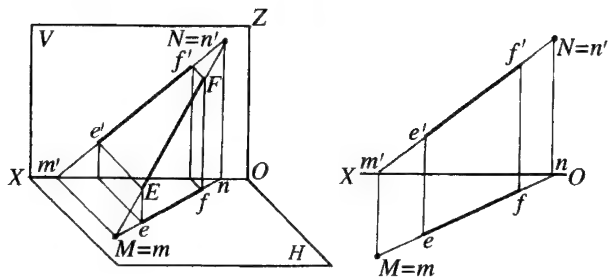


图 1　　　　图 2

投影)，即 m，n 在 ef 上，m'，n' 在 $e'f'$ 上.

2. 因迹点是投影面上的点，所以迹点的一个投影必在投影轴上(图中，m'，n 在轴上)，而另一个投影必与迹点本身重合(图中，m，n' 分别与 M，N 重合).

正面迹点(frontal trace point of straight line) 见“直线的迹点”.

水平迹点(horizontal trace point of straight line) 见“直线的迹点”.

侧面迹点(profile trace point of straight line) 见“直线的迹点”.

平面表示法(representation of plane) 确定平面位置的方法. 平面可用下列任一组几何元素的投影确定其空间的位置(如图所示)：

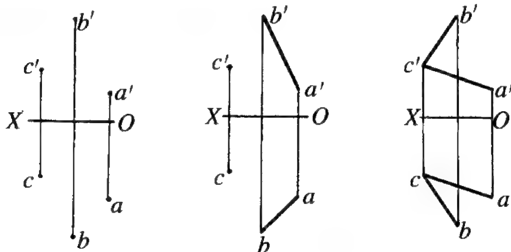


图 1　　　　图 2

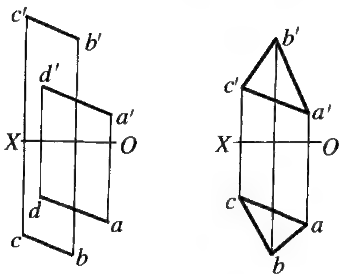


图 3　　　　图 4

1. 不在同一直线上的三点.

2. 一直线和线外一点.

3. 相交两直线.

4. 平行两直线.

5. 任意平面图形(三角形、圆或其他).

平面在空间的位置亦可用平面的迹线表示.

平面迹线(trace line of plane) 平面上的特殊直线. 平面与投影面的交线称为平面迹线. 如图 1、图 2 所示. 平面与 V 面的交线称为正面迹线，与 H 面的交线称为水平迹线，与 W 面的交线称为侧面迹线. 图中，分别以 Q_V，Q_H，Q_W 表示平面 Q 的三条迹线. 平面的迹线，两两相交，其交点在投影轴上，Q 平

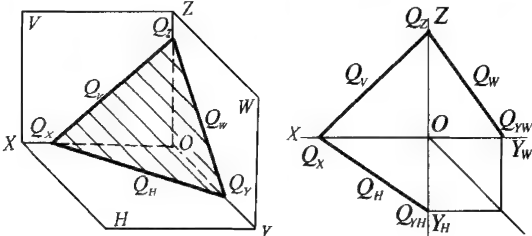


图 1　　　　图 2

面与三投影轴的交点分别用 Q_X，Q_Y，Q_Z 表示.

平面迹线是投影面上的直线，所以它的一个投影与其本身重合，另外两个投影与相应的投影轴重合. 如图中的 Q_H，其水平投影和它本身重合，正面投影和侧面投影分别和 OX 轴、OY 轴重合. 在投影图上表示迹线，一般只将迹线与自身重合的那个投影画出，并用符号标注，而和投影轴重合的投影不加标注.

正面迹线(frontal trace line) 见“平面迹线”.

水平迹线(horizontal trace line) 见“平面迹线”.

侧面迹线(profile trace line) 见“平面迹线”.

圆的投影(projection of circle) 一种最基本、常见的投影. 指不同位置圆在投影面上的反映. 若圆所在平面平行某一投影面，则圆在该投影面上的投影反映圆的实形，另外两投影积聚为线段. 图 1 为水平面上圆的投影. 如果圆所在平面垂直于某一投影面，则圆在该投影面上的投影积聚为线段，另外两个投影为椭圆. 如图所示为正垂面上圆的投影；如图 1，2，3.

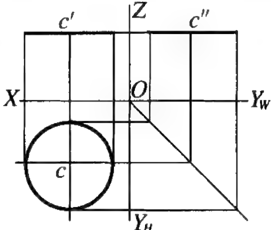


图 1

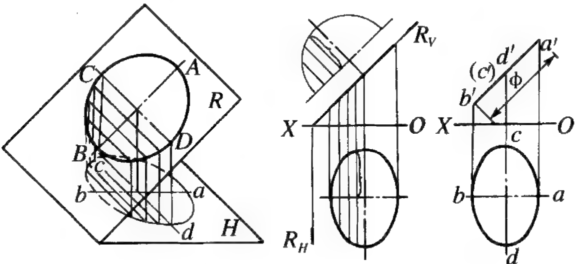


图 2　　　　图 3　　　　图 4

投影面平行面(parallel plane of projection plane)　一类具有特殊位置的平面. 指平行某一投影面,而与另外两个投影面垂直的平面.

正平面(frontal plane)　一种投影面平行面. 平行于 V 面的平面称为正平面. 如图 1、图 2 所示. 其投影特性为:

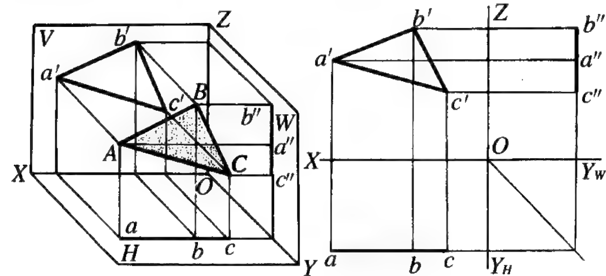


图 1　　　　图 2

1. 正平面的正面投影反映平面的实形;

2. 正平面的水平投影和侧面投影各为一直线,具有积聚性;

3. 正平面的水平投影平行于 OX 轴,即 $abc // OX$,正平面的侧面投影平行于 OZ 轴,即 $b''a''c'' // OZ$.

水平面(horizontal plane)　一种投影面平行面. 平行于 H 面的平面称为水平面. 如图所示. 其投影特性为:

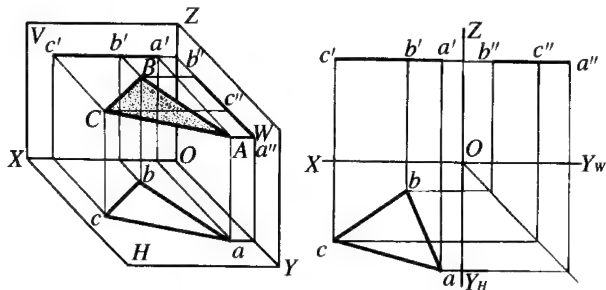


图 1　　　　图 2

1. 水平面的水平投影反映平面的实形.

2. 水平面的正面投影和侧面投影各为一直线,具有积聚性.

3. 水平面的正面投影平行于 OX 轴,即 $a'b'c' // OX$,水平面的侧面投影平行于 OY_W 轴,即 $a''c''b'' // OY_W$.

侧平面(profile plane)　一种投影面平行面. 平行于 W 面的平面称为侧平面. 如图所示. 其投影特性为:

1. 侧平面的侧面投影反映平面的实形.

2. 侧平面的正面投影和水平投影各为一直线,具有积聚性.

3. 侧平面的正面投影平行于 OZ 轴,即 $a'c'b' // OZ$,侧平面的水平投影平行于 OY_H 轴,即 $bac // OY_H$.

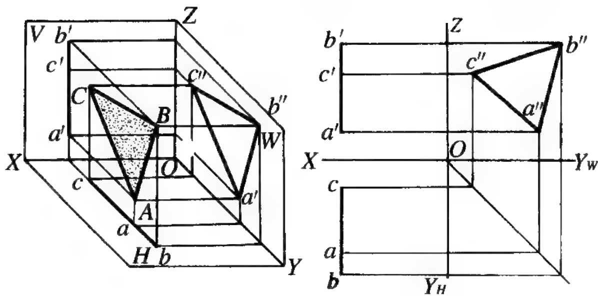


图 1　　　　图 2

投影面垂直面(perpendicular plane of projection plane)　一类具有特殊位置的平面. 指垂直某一投影面,而与另外两个投影面倾斜的平面.

正垂面(plane inclined in elevation)　一种投影面垂直面. 垂直于 V 面的平面称为正垂面. 如图所示. 其投影特性为:

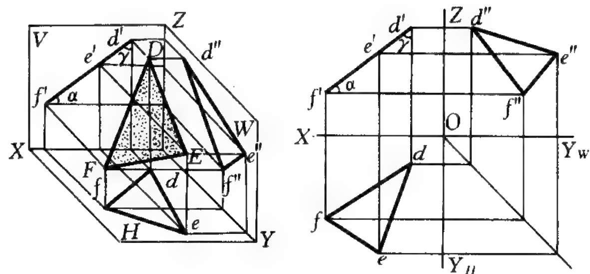


图 1　　　　图 2

1. 正垂面的正面投影为一直线,具有积聚性.

2. 正垂面的水平投影和侧面投影,不反映该平面的实形,而成为类似形.

3. 正垂面的正面投影与 OX 轴的夹角,反映该平面对 H 面的倾角 α,与 OZ 轴的夹角,反映该平面对 W 面的倾角 γ.

铅垂面(vertical plane)　一种投影面垂直面. 垂直于 H 面的平面称为铅垂面. 如图所示. 其投影特性为:

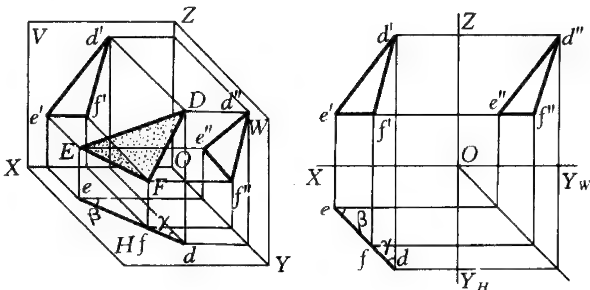


图 1　　　　图 2

1. 铅垂面的水平投影为一直线,具有积聚性.

2. 铅垂面的正面投影和侧面投影,不反映该平

面的实形，而成为类似形.

3. 铅垂面的水平投影与 OX 轴的夹角，反映该平面对 V 面的倾角 β，与 OY_H 轴的夹角，反映该平面对 W 面的倾角 γ.

侧垂面(plane inclined in profile)　一种投影面垂直面. 垂直于 W 面的平面称为侧垂面. 如图所示. 其投影特性为：

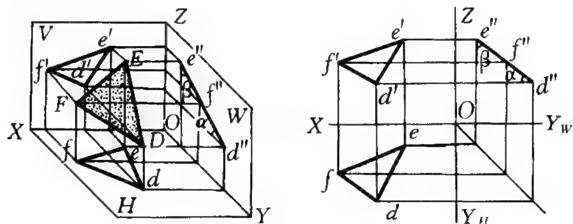


图 1　　图 2

1. 侧垂面的侧面投影为一直线，具有积聚性.

2. 侧垂面的正面投影和水平投影，不反映该平面的实形，而成为类似形.

3. 侧垂面的侧面投影与 OY_W 轴的夹角，反映该平面对 H 面的倾角 α，与 OZ 轴的夹角，反映该平面对 V 面的倾角 β.

一般位置平面(oblique plane)　一种平面. 即与三个投影面均无垂直、平行关系，而对三个投影面都倾斜的平面称为一般位置平面. 如图所示.

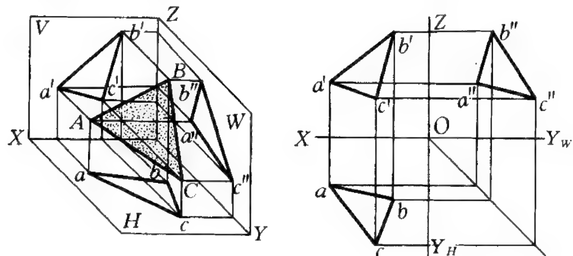
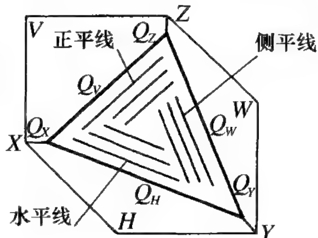


图 1　　图 2

一般位置平面和三个投影面既不平行，也不垂直，各个投影都不会成为直线，也不能反映空间平面图形的实形和对投影面的倾角，而在各个投影面上的投影都是空间平面图形的类似形.

平面上的投影面平行线(parallel line to projection plane on a plane)　平面上平行投影面的直线. 在一般位置平面上，可作正平线、水平线和侧平线. 平面上的投影面平行线其投影具有平面上直线的投影性质，也具有正平线、水平线和侧平线的投影特性.



平面上的正平线(frontal line on a plane)　平面上的一种特殊直线. 平面上的正平线是在平面上作出正平线. 如图所示. 其投影特性为：

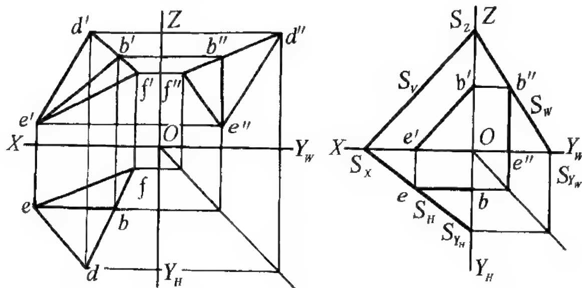


图 1　　图 2

1. 平面上正平线的水平投影平行 OX 轴.

2. 平面上正平线的正面投影反映该线实长，并且平行于平面的正面迹线 S_V.

平面上的水平线(horizontal line on a plane)　平面上的一种特殊直线. 平面上的水平线是在平面上做出水平线. 如图所示. 其投影特性为：

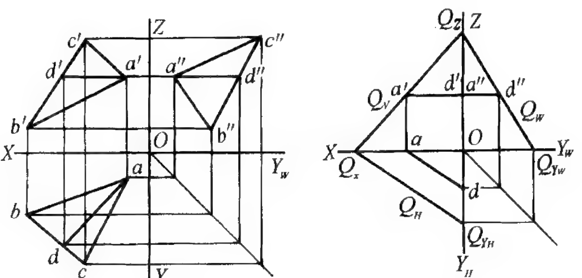


图 1　　图 2

1. 平面上水平线的正面投影平行 OX 轴.

2. 平面上水平线的水平投影反映该线实长，并且平行于平面的水平迹线 Q_H.

平面上的侧平线(profile line on a plane)　平面上的一种特殊直线. 平面上的侧平线是在平面上做出侧平线. 如图所示. 其投影特性为：

1. 平面上侧平线的正面投影平行 OZ 轴，水平投影平行 OY_H 轴.

2. 平面上侧平线的侧面投影反映该线实长，并且平行于平面的侧面迹线 S_W.

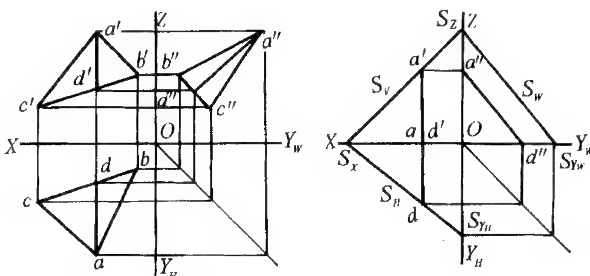
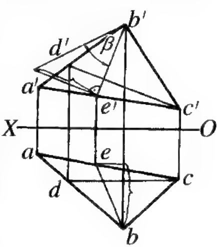


图 1　　图 2

平面上的最大倾斜线(maximum inclination line on a plane)　平面上的一种特殊直线. 指平面上垂直于该平面某一投影面平行线(也垂直于相应的迹线)的直线. 垂直于水平线的直线,称为对 H 面的最大倾斜线,垂直于正平线的直线,称为对 V 面的最大倾斜线,垂直于侧平线的直线,称为对 W 面的最大倾斜线. 平面上最大倾斜线与投影面的夹角就是该平面对投影面的倾角. 如图,就是利用平面△ABC 上最大倾斜线 BE 求出该平面对 V 面的倾角 β.



平面投影的积聚性(projection-coincidence nature of plane edge view)　投影法的基本概念之一. 当平面垂直某一投影面时,平面在该投影面上的投影为直线,此时,平面上的点、线、图形在该投影面上的投影均与此直线重合,此直线(平面的投影)称为有积聚性.

投影变换(projection transformation)　画法几何的基本概念之一. 它是一种解决几何元素的位置、实形以及度量等问题的重要方法. 研究如何改变空间几何元素对投影面的相对位置,以达到简化解决空间度量问题和定位问题.

辅助面法(auxiliary plane method)　一种解题方法. 所谓辅助面法,就是使空间几何元素的位置保持不变,以新的投影面更换原有的投影面,使空间几何元素与新投影面处于特殊位置,以达到解题的目的. 如图所示. 图中,△ABC 为一铅垂面,在 V 面和 H 面投影体系中,两个投影均不反映实形. 现取一平行于三角形且垂直于 H 面的 V_1 面来代替 V 面,则新的 V_1 面和 H 面构成一个新的两面体系 V_1/H. 三角形在新体系中 V_1 面上的投影△$a_1'b_1'c_1'$ 就反映△ABC 的实形. 再以 V_1 面和 H 面的交线 O_1X_1 为轴,使 V_1 面旋转到与 H 面重合,就得出 V_1/H 体系的投影. 新投影面必须符合以下两个条件:

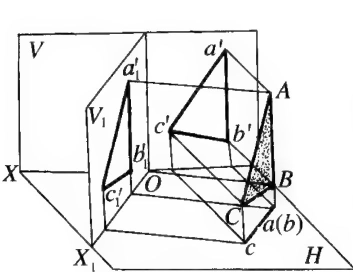


图 1

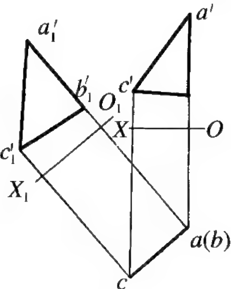


图 2

1. 新投影面必须垂直于原来的一个投影面.

2. 新投影面必须与空间元素处于有利于解题的位置.

重合法(coincidence method)　一种解题方法. 平面绕其一条迹线旋转,重合在相应的投影面上,这种方法称为重合法. 如图 1、图 2 所示. 在图中,平面 P 绕它的水平迹线 P_H 旋转,使平面 P 重合于 H 面时,在 H 面上即能反映 P 面上△ABC 的实形. 当平面绕着它的迹线旋转,使之重合到投影面上时,平面上点的运动轨迹是绕迹线旋转的圆弧,圆弧所在的平面,垂直于迹线,也垂直于迹线所在的投影面,圆弧的半径等于点到迹线的真实距离.

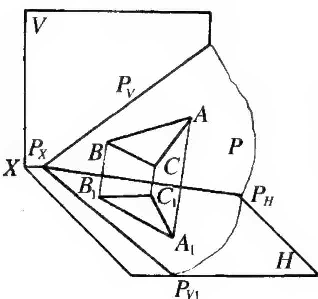


图 1

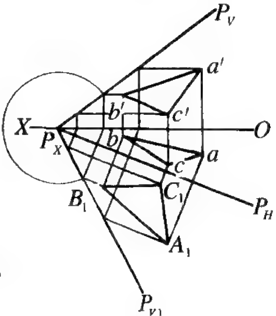


图 2

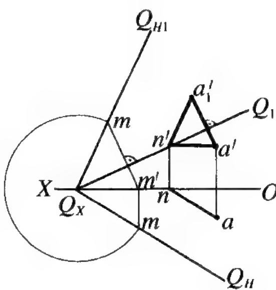


图 3

在图 3 中为 Q 平面上 A 点的重合作图. 其作法如下:

1. 先将 Q 平面绕 Q_V 重合到 V 面上,得出 Q_{H1}.

2. 过 A 点作水平线 NA (na, $n'a'$).

3. 过 n' 作直线 $n'a_1' /\!/ Q_{H1}$.

4. 过 a' 作直线垂直 Q_V,使其与直线 $n'a_1'$ 相交于 a_1' 点. a_1' 点即为 A 点重合后的位置.

旋转法(revolution method)　一种解题方法. 所谓旋转法,就是使投影面保持不动,将空间几何元素绕某一轴旋转,使其旋转到有利于解题的位置. 如图所示. 图中,△ABC 为一铅垂面,在 V 面和 H 面的投影,均不反映实形. 现使其绕垂直 H 面的轴

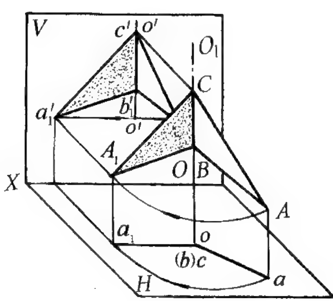


图 1

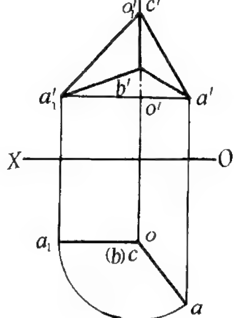


图 2

OO_1 旋转成为正平面$\triangle A_1BC$，此时$\triangle A_1BC$ 的正面投影$\triangle a_1'b'c'$反映$\triangle ABC$ 的实形.

曲线曲面

曲线(curve)　一种几何图形. 它是一个动点连续运动的轨迹. 当点的运动无一定的规律，则形成不规则曲线；如果点按一定的规律运动，则形成有规则的曲线，如圆、螺旋线等. 曲线的投影在一般情况下仍是曲线，为此只要做出曲线上一系列点的投影，并将各点的同面投影光滑地连结起来，即得到曲线的投影，如图 1 所示. 直线与曲线相切，其投影也必相切. 如图 2、图 3 中的 直线 Ⅰ Ⅱ 与曲线 $ABCD$ 相切于C点，则直线 12 切于$abcd$的点c，直线 $1'2'$ 切于 $a'b'c'd'$ 的点 c'.

图 1

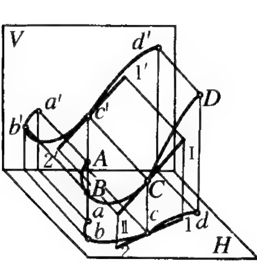


图 2

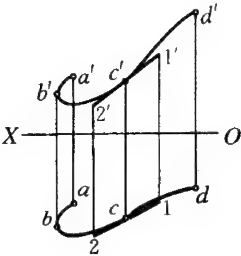


图 3

平面曲线(plane curve)　一类重要的曲线. 所谓平面曲线，是指曲线上所有的点都在同一平面内，如圆、椭圆、抛物线、双曲线、渐开线、摆线等. 平面曲线除了具有曲线的一 般投影性质外，还有下面的投影性质：平面曲线所在的平面平行于某一投影面时，则在该投影面上的投影反映曲线的实形，如图 1 所示. 平面曲线所在的平面垂直于某一投影面时，则在该投影面上的投影，积聚成一直线，如图 2 所示. 平面曲线的奇异点的投影，仍为曲线投影的同名奇异点，即拐点、尖点及两重点投影后仍为拐点、尖点及两重点，如图 3 所示.

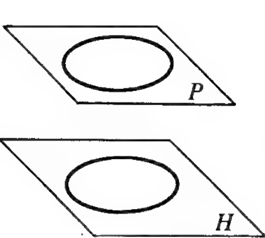


图 1

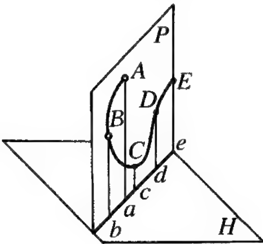


图 2

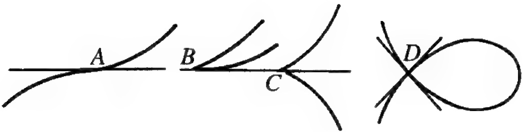


图 3

空间曲线(space curve)　一类重要的曲线. 所谓空间曲线，是指曲线上连续四点不在同一平面内，如螺旋线. 空间曲线的投影是曲线，有时亦会出现重影点. 如图所示的空间曲线，在空间并不相交，但点B,D的水平投影点b,d重合. 因此点b,d是曲线上的两个重影点B,D的投影.

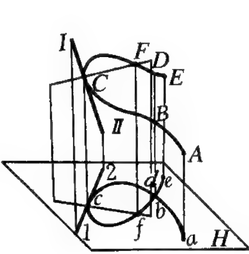


图 1

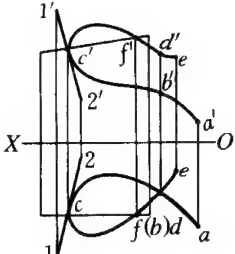
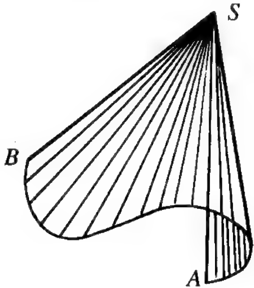


图 2

圆锥曲线(conic section)　一类常见的平面曲线. 是平面曲线中的圆、椭圆、抛物线和双曲线等四种曲线的总称，用不同位置的平面截圆锥可得这四种曲线. 圆锥曲线投影后，一般性质不变，即抛物线的投影为抛物线；双曲线的投影为双曲线；椭圆的投影为椭圆.

曲面(curved surface)　一种几何图形. 它是一条动线(直线或曲线)在空间连续运动的轨迹. 产生曲面的那条线(直线或曲线)称为母线. 控制母线运动的线或面，称为导线或导面. 母线在曲面上任一位置，称为素线. 如图，曲线 AB 为导线，直线 SB 为母线，SB 运动的各个位置为素线.



母线(generatrix)　见"曲面".

导线(directrix)　见"曲面".

素线(element)　见"曲面".

直线面(straight-line curved surface)　一类基本、常见的曲面. 由直母线形成的曲面称为直线面. 如柱面、圆锥面和正螺旋面等.

单曲面(single curved surface)　一种直线面. 单曲面的连续两素线彼此平行或相交，位于同一平面内，因而是可展曲面. 所以，单曲面又称为可展直线面. 工程上常见的有柱面、锥面及切线曲面.

可展直线面(developable straight-line curved surface) 即"单曲面".

柱面(cylinder) 一种常见的单曲面.即直母线

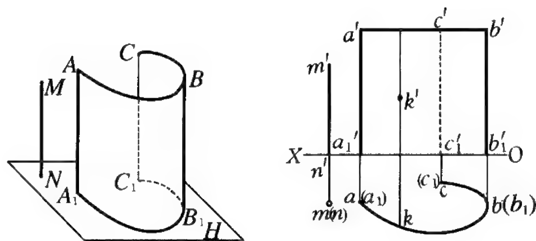


图 1　　　　图 2

沿曲导线移动,并平行于另一直导线,就形成了柱面.如图 1,直母线 AA_1 沿曲线 $A_1B_1C_1$ 移动,且始终平行于直导线 MN,所得曲面是柱面,上述曲导线可以是闭合的,也可以是不闭合的.其投影图的画法,如图 2 所示.假如柱面上有一 K 点,则 K 一定在柱面的一条素线上.当曲导线为圆,直导线垂直于圆

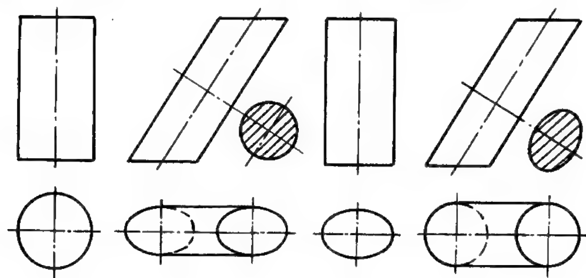
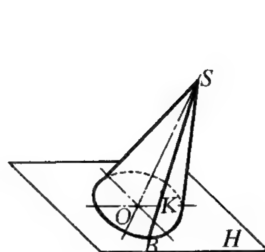


图 3　　图 4　　图 5　　图 6

平面时称为圆柱面.圆柱面的轴线垂直于柱底时 称为正圆柱面,如图 3 所示.圆柱面的轴线倾斜于柱底时称为斜圆柱面,如图 4 所示.当曲导线为椭圆,直导线垂直椭圆平面时,称为椭圆柱面.椭圆柱面的轴线垂直于柱底时,称为正椭圆柱面,如图 5 所示.椭圆柱面的轴线倾斜柱底时,称为斜椭圆柱面,如图 6 所示.圆柱面与椭圆柱面的区别是:垂直于柱面轴线的截平面(正截面)在柱面上截出的是圆,则此柱面就是圆柱面,若截出的是椭圆,则此柱面就是椭圆柱面.柱面各对称平面的交线,称为柱面的轴线,柱面可以是有轴线的或无轴线的.

锥面(cone) 一种常见的单曲面.即直母线沿曲导线移动,并始终通过一定点(锥顶),就形成了锥面.如图 1,点 S 为锥顶,曲导线为圆,SB 为母线,SB 运动的各个位置是素线,所得曲面是锥面,上述曲导线可以是闭合的,也可以是不闭合的.其投影图的画法,如图 2 所示.假如锥面上有一 K 点,则 K 点一定在过锥顶 S 的素线上.当截平面与锥面的轴线垂直相交,其截交线为圆,则此锥面是圆锥面,当圆锥面轴线垂直于锥底时,称为正圆锥面,如图 3 所示.当轴线倾斜于锥底时,则称为斜圆锥面,如图 4



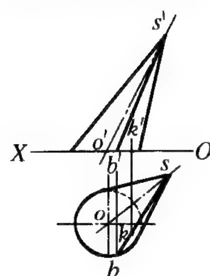


图 1　　　　图 2

所示.当截平面与锥面的轴线垂直相交,其截交线为椭圆,则此锥面为椭圆锥面,当椭圆锥面轴线垂直于锥底时,称为正椭圆锥面,如图 5 所示.当椭圆锥面

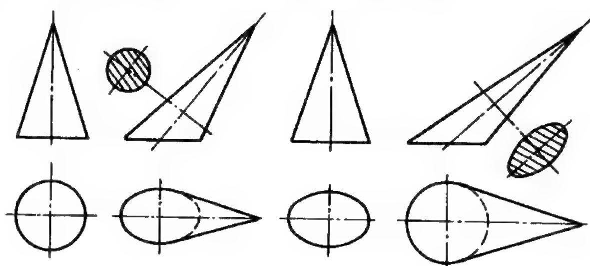
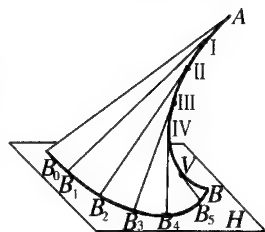


图 3　　图 4　　图 5　　图 6

轴线倾斜于锥底时,称为斜椭圆锥面,如图 6 所示.锥面各对称平面的交线,称为锥面的轴线.锥面可以是有轴线的或无轴线的.

切线曲面(tangent surface) 一种特殊的单曲面.即直母线始终切于一空间曲导线,其运动轨迹就形成了切线曲面.如图所示,直母线 AB_0 切一空间曲导线 AB 于点 A,直母线连续运动到素线 I B_1 位置,切空间曲导线 AB 于点 I,直母线如此连续运动,并始终保持与曲导线相切所形成的曲面.



扭曲面(warped surface) 一种特殊的直线面.扭曲面的连续两素线彼此交叉,不在同一平面内,因而是不可展曲面.所以,扭曲面又称为"不可展直线面".工程上常见的有双曲抛物面、锥状面、柱状面、斜螺旋面.

不可展直线面(non-developable straight-line curved surface) 见"扭曲面".

柱状面(cylindroid) 一种扭曲面.即一直母线沿着两条曲导线移动,并始终与一导平面保持平行,就形成了柱状面.如图 1 所示,在图中其曲导线为 $ABCDE$ 和 $FGIKL$,导平面为 H.其投影图如图 2 所示.为了明确表示出曲面的投影,应画出足够数量的素线的投影,如 af 和 $a'f'$,bg 和 $b'g'$等.同时还

图 1

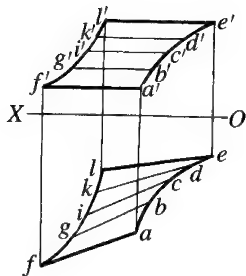


图 2

可以看出 $a'f' \parallel b'g' \parallel c'i' \parallel d'k' \parallel e'l' \parallel OX$（因母线移动时与 H 面保持平行，各条素线都应是水平线），而它们的水平投影并不平行，如果平行就成了柱面．柱状面的相邻两素线为交叉，所以是不可展曲面．

锥状面(conoid)　一种扭曲面．即一直母线沿着两条导线（一条为直线，另一条为曲线）移动，并始

图 1

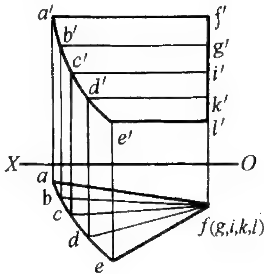


图 2

终与一导平面平行，就形成了锥状面．如图 1 所示，在图中其曲导线为 $ABCDE$，直导线为 FL，导平面为 H．其投影图如图 2 所示，直导线 FL 垂直于 H 面，H 面是导平面．如果直导线变为一个点，则锥状面就变成锥面．锥状面相邻两素线彼此交叉，所以是不可展曲面．

双曲抛物面(hyperbolic paraboloid)　一种扭曲面．即一直母线沿着两条直导线（导线为交叉两直线）移动，并始终与一导平面平行，就形成了双曲抛物面．如图 1 所示，AD，BC 两交叉直线为导线，H 面为导平面，母线 AB 沿 AD，BC 移动，并始终平行于 H 面．图 2 所示为双曲抛物面的投影图，图中铅垂面 R（水平迹线为 R_H）为导平面，AD，BC 为导线（交叉两直线），母线 CD 沿两导线移动，并始终平行于导平面 R．此双曲抛物面也可看做：以铅垂面 Q（水平迹线为 Q_H）为导平面，BA，CD 为导线（交叉两直线），母线 AD 沿两导线移动，始终平行于导平面 Q 而形成．

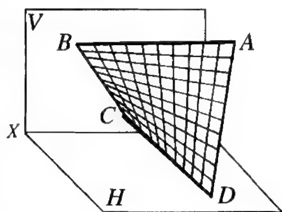


图 1

上述两组素线的水平素线的交点 O，称双曲抛物面的顶点．平行导平面并通过 O 的铅垂线，称双曲抛物面的轴．任何平行于轴 O 而不平行导平面的平面，与此曲面的交线为抛物线；不平行于轴的平面与此曲面的交线为双曲线．此曲面故名双曲抛物面．平行于导平面 R 或导平面 Q 的平面，与此曲面的交线为直线，通过双曲抛物面顶点 O 而垂直于轴的平面，与曲面交于两素线．

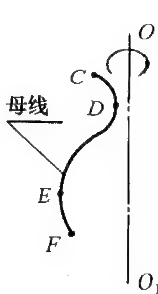
图 2

曲线面(double-curved surface)　一类曲面．由曲母线形成的曲面称为曲线面，如球面、圆环面等．曲线面都是不可展曲面．

定线曲面(constant-curve surface)　一类曲线面．定线曲面的曲母线在运动过程中，形状和大小保持不变时形成的曲面称为定线曲面，如曲线回转面．

变线曲面(varied-curve surface)　一类曲线面．变线曲面的曲母线在运动过程中，形状和大小不断变化，此时形成的曲面称为变线曲面．属于变线曲面的，如汽车车身、船舶壳体、飞机机身、泵体流道、推进器叶片等所用的曲面．

回转面(surface of revolution)　一类常见的曲线面．母线（直线、曲线）绕一轴线回转形成的曲面称为回转面．如图 1 所示，其母线是几段圆弧组成，并绕轴 OO_1 旋转一周所形成的曲面称为回转面．其投影图如图 2 所示．



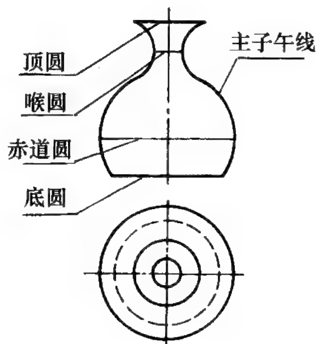


图 1　　图 2

子午线(meridian line)　回转面上的一种曲线．曲线回转面上的任一素线，就是一条子午线，其中形成回转面轮廓的素线称为主子午线．

主子午线(principal meridian line)　见“子午线”．

纬圆(latitudinal circle)　回转面上的一种圆．回转面上任一点的轨迹是圆，这些圆通常称为纬圆，

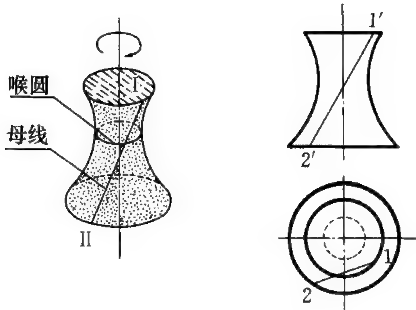
也称为纬线. 如"回转面"条目图 1 中 C, D, E, F 等点所形成的圆(参见"回转面").

赤道圆(equator circle) 一种特殊的纬圆. 回转面上比相邻两侧的纬圆都大的纬圆称为赤道圆,如"回转面"条目图 1 中 E 点形成的圆(参见"回转面"). 由于回转面形状不同,有的回转面可能有几个赤道圆.

喉圆(throat circle) 一种特殊的纬圆. 回转面上比相邻两侧的纬圆都小的纬圆称为喉圆,如"回转面"条目图 1 中 D 点形成的圆(参见"回转面"). 由于回转面形状不同,有的回转面可能有几个喉圆.

直线回转面(straight-line surface of revolution) 一类回转面. 一直线作母线,另一直线作轴线,母线绕轴旋转形成的曲面称为直线回转面,如圆柱面、圆锥面、单叶双曲回转面.

单叶双曲回转面(one sheet hyperboloid revolution surface) 一种直线回转面. 即一直母线绕与



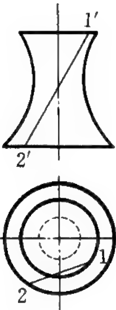
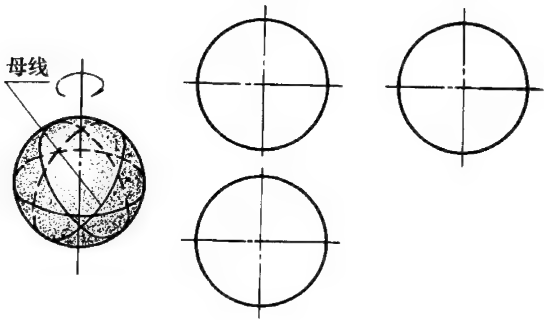


图 1　　图 2

其交叉的回转轴回转后形成的曲面,如图 1 所示. 用包含轴线的平面截切单叶双曲回转面,其截交线为双曲线. 因此,单叶双曲回转面也可看做是以双曲线为母线,绕它的虚轴回转而成. 其投影图如图 2 所示,其 V 面投影的轮廓线为双曲线,在 H 面表示了喉圆和底圆的投影.

曲线回转面(curved-line surface of revolution) 一类曲线面. 任意的平面曲线绕与其在同一平面的回转轴线旋转形成的曲面称为曲线回转面,如球面、环面等.

球面(spherical surface) 一种常见的曲线回转面. 即一圆母线以它本身的一条直径作为轴线,回转而成的曲线回转面称为球面,如图 1 所示. 其投影图如图 2 所示. 圆球的三个投影均为圆,且其直径等于球直径.



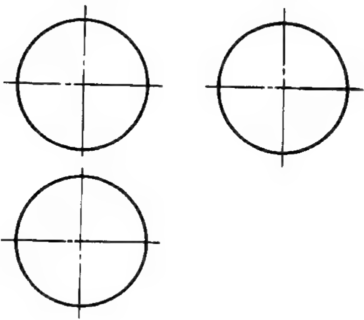
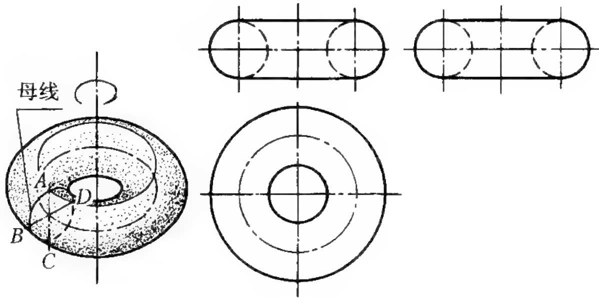


图 1　　图 2

环面(torus) 一种常见的曲线回转面. 一圆母线绕与它在同一平面内且不通过圆心的一条轴线回转而成的曲线回转面称为环面,如图 1 所示. 母线 ABC 半个圆回转成外环面,ADC 半个圆回转成内环面,内、外环面的分界线是两个纬圆,称为分界圆.



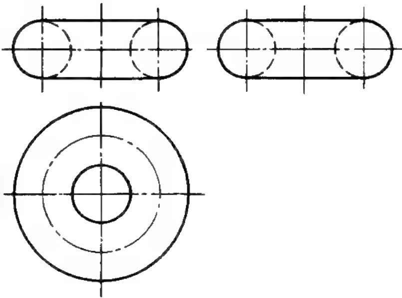
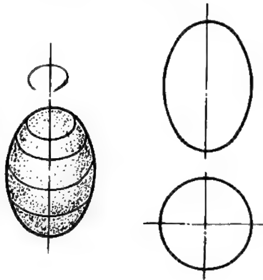


图 1　　图 2

其投影图如图 2 所示.

分界圆(dividing circle) 见"环面".

回转椭球面(ellipsoid surface of revolution) 一种曲线回转面. 以椭圆为母线,绕它本身的一条轴(长轴或短轴)回转而成的曲线回转面称为回转椭球面. 图 1 所示的回转椭球面是绕长轴旋转形成的,V 面投影的轮廓线是椭圆本身的大小,而 H 面投影的轮廓线是以短轴为直径的圆,如图 2 所示.



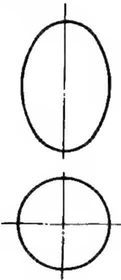
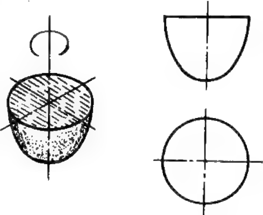


图 1　　图 2

回转抛物面(paraboloid surface of revolution) 一种曲线回转面. 以抛物线为母线,绕它本身的轴回转而成的曲线回转面称为回转抛物面,如图 1 所示. 其投影图如图 2 所示,V 面投影的轮廓线就是抛物线本身,而 H 面投影的轮廓线是圆.



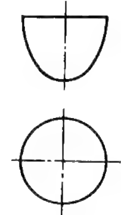


图 1　　图 2

圆柱螺旋线(cylindrical helix) 一种常见的曲线. 一个动点沿圆柱面的母线作匀速直线运动,同时该母线又绕圆柱面的轴线作匀速转动,点的这种复合运动的轨迹称为圆柱螺旋线,如图所示. 螺旋线可分为左旋和右旋两种,图 1 所示为右旋螺旋线,而图

2 所示为左旋螺旋线. 母线旋转一周动点在母线上移动的一段距离为导程 H. 决定一圆柱螺旋线的三个基本要素是:圆柱螺旋线的直径 D、导程 H 及旋向. 根据已知三要素 D, H,旋向,就可以画出螺旋线的投影图,如图 3 所示. 如将圆柱表面展开,则螺旋线展成一直线,这条直线为直角三角形的斜边,该直角三角形的一条直角边长为导程 H,另一直角边长为圆柱底圆的周长 πD. 显然,在一个导程内,螺旋线的长度为 $\sqrt{(\pi D)^2+H^2}$. 该直角三角形的锐角 λ 称为螺旋线的升角. 其中,

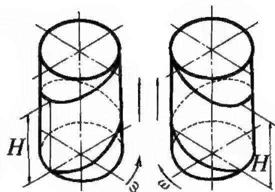


图 1　　图 2

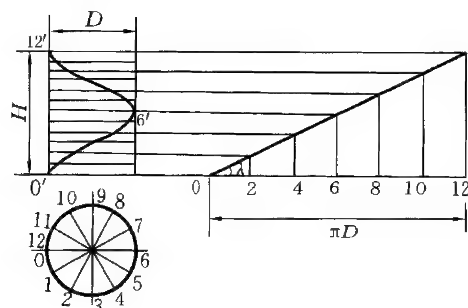
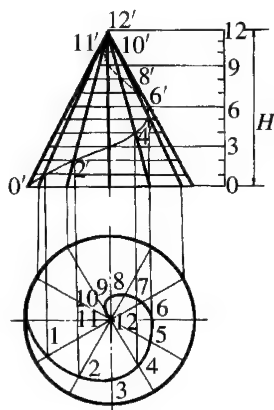


图 3

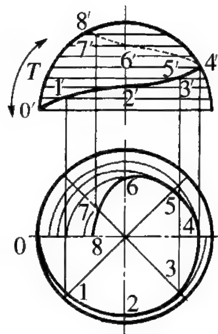
$$\lambda = \arctan \frac{H}{\pi D}.$$

螺旋线的升角(helix angle)　见"圆柱螺旋线".

圆锥螺旋线(conical helix)　一种常见的曲线. 当一动点沿正圆锥面的母线作匀速直线运动,同时该母线沿圆锥轴线作匀速转动,点的这种复合运动的轨迹称为圆锥螺旋线. 母线回转一周,其动点沿圆锥轴线方向移动的距离称为导程 H. 图示为圆锥螺旋线的投影图. 圆锥螺旋线的 H 面投影为一阿基米德螺线. 圆锥螺旋线导程也可沿素线度量. 圆锥螺旋线有左旋与右旋两种.



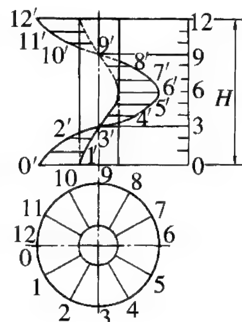
球面螺旋线(spherical helix)　一种常见的曲线. 动点沿球面的子午线作等速圆弧运动,同时该子午线绕其轴线作等角速回转运动,动点的轨迹称为球面螺旋线,如图所示. 球面螺旋线的导程 T 沿子午线度量. 有左旋和右旋两种.



回转面螺旋线(helix of surface of revolution)　一种常见的曲线. 动点沿回转面的子午线作等速圆弧运动,同时该子午线绕轴作等角速回转运动,该点的轨迹称为回转面螺旋线. 回转面螺旋线的导程沿子午线度量. 球面螺旋线即是回转面螺旋线的一种. 有左旋和右旋两种.

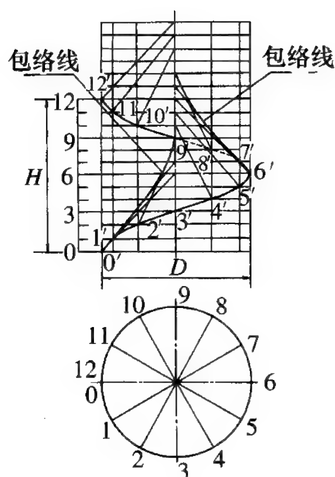
螺旋面(helicoidal surface)　一类常见的曲面. 以螺旋线和它的轴线为导线,直母线(也可以是曲母线)沿两条导线滑动,并始终与轴线交成定角所形成的曲面称为螺旋面. 同螺旋线一样,螺旋面也分成左旋和右旋两种. 在形成螺旋面的过程中,母线上各点轨迹都是螺旋线. 这些螺旋线导程相等. 画出螺旋线和轴线的投影,再画出若干直素线的投影以及包络线,就得到螺旋面的投影.

正螺旋面(right helicoid)　一种常见的螺旋面. 直母线沿一条圆柱螺旋线运动,并始终与其轴线垂直相交所形成的曲面称为正螺旋面. 从正螺旋面的形成看,正螺旋面也可以说是一种锥状面. 如图所示,它是由直线段沿圆柱螺旋线作螺旋运动而形成的. 螺旋线的导程为 H,方向为右旋. 因螺旋线的轴线为铅垂线,所以正螺旋面的所有素线都是水平线,该螺旋面投影作法是:画出两导圆柱的右旋螺旋线;连结两螺旋线对应点得螺旋面相应素线的投影;最后判别可见性. 将不可见的螺旋线、素线和圆柱轮廓线画成虚线. 图示即为圆柱正螺旋面. 若直母线沿一圆锥螺旋线运动,并始终与圆锥的轴线垂直相交,所形成的曲面称为圆锥螺旋面.



圆锥螺旋面(conical helicoid)　见"正螺旋面".

斜螺旋面(oblique helicoid)　一种常见的螺旋面. 以任意回转面螺旋线为导线,当直母线运动时始终与回转面轴线成定角斜交所形成的曲面称为斜螺旋面. 若直母线沿一条圆柱螺旋线运动并始终与其轴线成定角斜交,则形成圆柱斜螺旋面. 如图所示为一右旋圆柱斜螺旋面. 设螺旋线的轴线为铅垂线,直素线与轴线交成定角,螺旋线导圆柱直径为 D,导程为 H,右旋. 该螺旋面投影作法是:画出辅助圆柱面的两投影;在水平投影上,以 o 为起点,将圆分成等



分，作为斜螺旋面各素线的水平投影；在正面投影上，从 o' 作一直线与轴线成已知直素线与轴的定角，即为斜螺旋面一条素线的正面投影. 如此做出一系列素线，判别可见性，再画出外形轮廓所需的包络线.

圆柱斜螺旋面(oblique cylindrical helicoid) 见“斜螺旋面”.

立体及其截断与相贯

平面立体(plane-surface solid) 一类常见的立体. 表面由平面包围而成的立体称为平面立体. 常见的平面立体有棱柱和棱锥. 如图 1 至图 4 所示. 图 1、图 2 中，棱柱是由棱面和顶面、底面所围成，相邻

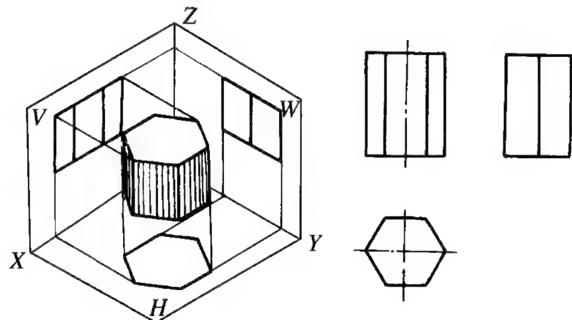


图 1　　　图 2

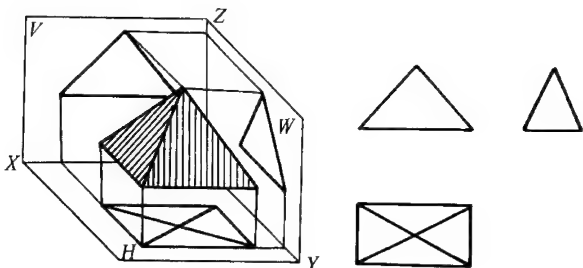


图 3　　　图 4

两棱面的交线，称为棱线. 图 3、图 4 中，棱锥是由棱面和底面所围成，各棱面是有一个公共顶点的三角形.

曲面立体(curved-surface solid) 一类常见的立体. 表面由曲面或由曲面和平面包围而成的立体称为曲面立体. 常见的曲面立体有圆柱、圆锥、圆球、圆环. 由于包围上述这些立体的曲面都属于回转曲面，故这些立体又称为回转体.

回转体(revolving body) 见“曲面立体”.

圆柱(cylinder) 一种常见的曲面立体. 由圆柱面和顶面、底面所围成的立体称为圆柱. 图 1、图 2

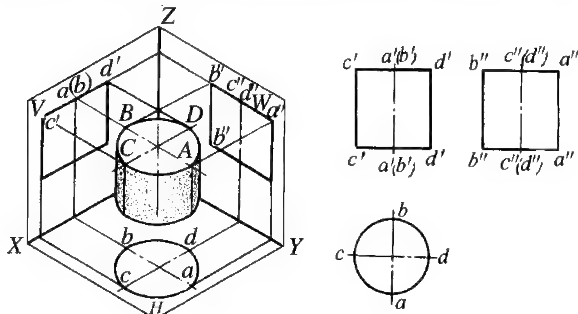


图 1　　　图 2

分别为轴线垂直于 H 面的圆柱立体图和投影图. 由于圆柱的轴线为铅垂线，因此圆柱的水平投影积聚为一圆，圆柱顶面、底面的水平投影与此圆重合. 圆柱的正面投影和侧面投影是大小相同的矩形，其上、下边分别是顶面、底面的投影，左、右两边是圆柱轮廓线的投影.

圆锥(cone) 一种常见的曲面立体. 由圆锥面和底面所围成的立体称为圆锥. 图 1、图 2 为轴线垂

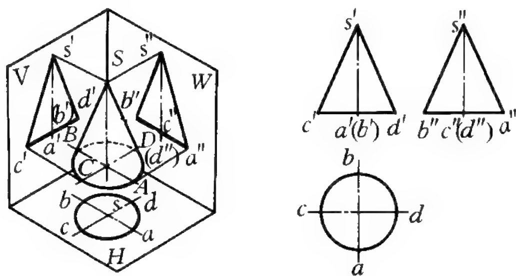


图 1　　　图 2

直于 H 面圆锥的立体图和投影图. 圆锥的水平投影为圆，此圆是圆锥面的水平投影，也是底面的水平投影. 正面投影和侧面投影是大小相同的等腰三角形，其底边是底面的投影，两腰是圆锥轮廓线的投影.

圆球(sphere) 一种常见的曲面立体. 由球面所围成的立体称为圆球. 图 1、图 2 为球的立体图和投影图. 球的三个投影都是大小相同的圆，圆的直径等于球的直径.

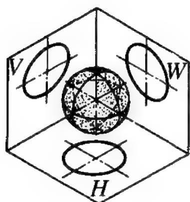


图 1

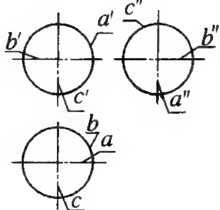


图 2

截交线(section)　一种直线. 指平面与立体表面相交的交线. 该平面称为截平面. 截交线所围成的

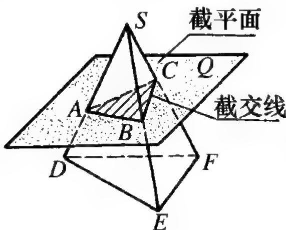


图 1

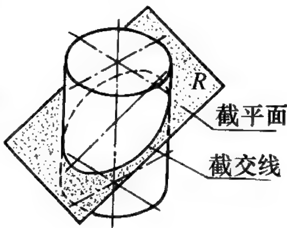


图 2

平面图形称为截断面. 如图 1、图 2 所示. 截交线的形状取决于基本几何体的形状和截平面与基本几何体的相对位置.

截交线的特性为:

1. 截交线上的点既在立体表面上, 又在截平面上, 是立体与截平面的公有点.

2. 由于立体都有一定范围, 故截交线是封闭的平面图形.

截平面(cutting plane)　见"截交线".

截断面(section)　见"截交线".

圆环(torus)　一种常见的曲面立体. 由环面所围成的立体称为圆环. 图 1、图 2 分别为轴线垂直于 H 面圆环的立体图和投影图, 是由一个圆围绕与圆共面, 但位于此圆外的轴线旋转而成. 外半圆旋转所成的曲面称为外环面, 内半圆旋转所成的曲面称为内环面.

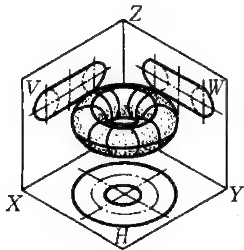


图 1

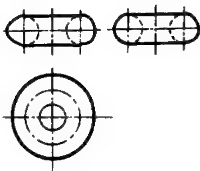


图 2

圆柱截交线(cylindrical section)　一种常见的截交线. 平面与圆柱相交, 其交线称为圆柱截交线. 由于截平面与圆柱轴线的相对位置不同, 其截交线的形状也不同. 当截平面平行圆柱轴线时, 截交线为矩形; 当截平面垂直圆柱轴线时, 截交线为圆; 当截平面倾斜圆柱轴线时, 截交线为椭圆. 如表所示.

截平面的位置	平行于轴线	垂直于轴线	倾斜于轴线
截交线的形状	矩　形	圆	椭　圆
立体图			
投影图			

圆锥截交线(conic section)　一种常见的截交线. 平面与圆锥相交, 其交线称为圆锥截交线. 由于截平面与圆锥轴线的相对位置不同, 其截交线的形

截平面位置	过锥顶	不过锥顶			
		θ=90°	θ>α	θ=α	θ<α
截交线形状	三角形	圆	椭圆	抛物线	双曲线
立体图					
投影图					

状也不同. 如表所示, 当截平面通过锥顶且倾斜圆锥轴线时, 截交线为三角形; 当截平面不过圆锥顶点且垂直于圆锥轴线($\theta=90°$)时, 截交线为圆; 当截平面不过圆锥顶点且倾斜圆锥轴线($\theta>\alpha$)时, 截交线为椭圆; 当截平面不过圆锥顶点且平行圆锥素线($\theta=\alpha$)时, 截交线为抛物线; 当截平面不过圆锥顶点且倾斜($\theta<\alpha$)或截平面平行圆锥轴线时, 截交线为双曲线.

圆球截交线(section of sphere)　一种常见的截交线. 平面与圆球相交, 其交线称为圆球截交线.

截平面平行于投影面，截交线在该投影面的投影为圆(俯视)	截平面倾斜于投影面，截交线在该投影面的投影为椭圆(俯、左视)
	D

不论截平面处于何种位置, 它们的截交线总是圆. 但当截平面位置不同时, 截交线的投影也不一样, 当截平面与投影面平行时, 其截交线的两个投影为直线, 一个投影为圆; 当截平面与投影面倾斜(垂直某一投影面)时, 其截交线两个投影为椭圆, 一个投影为直

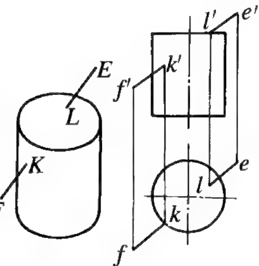
线. 如上表所示.

圆环截交线(section of torus)　一种常见的截交线. 平面与圆环相交,其交线称为圆环截交线. 由于截平面与圆环的相对位置不同,其截交线的形状

截平面垂直于轴线时，截交线为两个同心圆	截平面通过轴线时，截交线为两个圆	截平面平行于轴线时，截交线是平面曲线（其形状随截平面到轴线的距离不同而改变）

也不同. 当截平面垂直圆环轴线时,截交线是两个同心圆;当截平面通过圆环轴线时,截交线是两个圆;当截平面平行圆环轴线时,截交线为平面曲线,但由于截平面距圆环轴线远近不同,截交线形状也不同. 如上表所示.

贯穿点(piercing points)　画法几何的基本概念之一. 直线与立体表面的交点称为贯穿点. 贯穿点既在立体表面上,又在直线上,它总是成双存在,即有一个穿入点和一个穿出点. 如图所示.



相贯线(intersection of solid objects)　画法几何的基本概念之一. 两立体表面相交的交线称为相贯线. 两立体相交可分为三种情况:

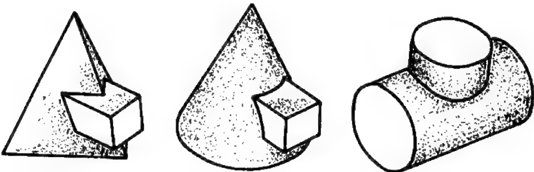


图 1　　图 2　　图 3

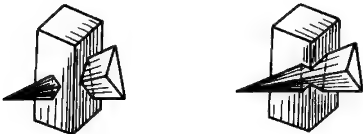


图 4　　图 5

1. 平面立体与平面立体相交,如图 1 所示.

2. 平面立体与曲面立体相交,如图 2 所示.

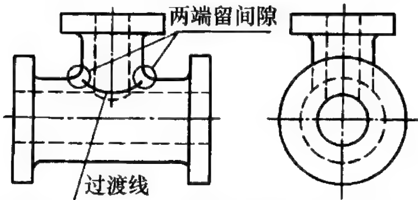
3. 曲面立体与曲面立体相交,如图 3 所示.

两立体相交,根据两立体相对位置不同,又可分为全交与互交两种,图 4 为全交;图 5 为互交. 相贯线的基本特性是:

1. 相贯线上的点是两立体表面的公共点.

2. 由于立体有一定范围,故相贯线是闭合的.

过渡线(transition line)　机件表达的基本概念之一. 它表示机械零件表面一种构造的画法. 机械零件,特别是铸件,常常在两立体表面相交处由一个曲



面圆滑地连结起来,有了圆滑过渡部分,相贯线的明显程度大为降低,这种不太明显的相贯线称为过渡线. 如图所示.

展　开　图

展开图(development)　一种平面图形. 将立体表面的实际形状和大小,依次展平在一个平面上的过程称为立体表面展开,所得的图形称为展开图.

可展曲面展开图(development of developable curved surface)　展开图的一种. 曲面上的两条连

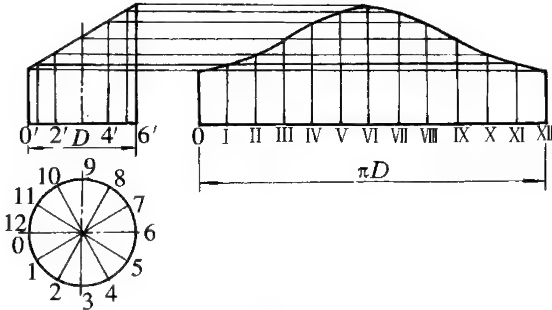


图 1

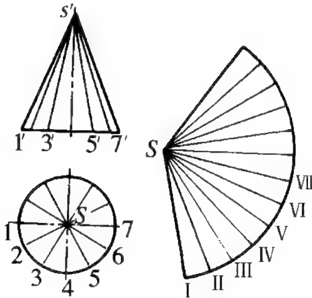
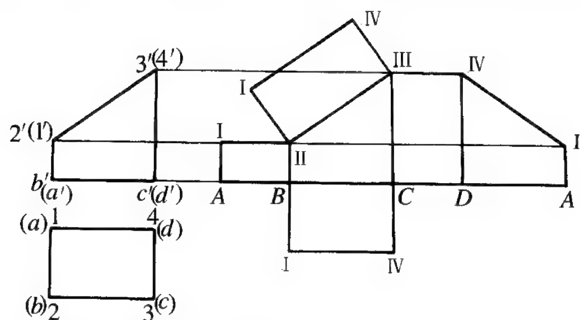


图 2

续素线平行或相交时,该曲面称为可展曲面,如圆柱、圆锥、切线曲面等. 图 1 为截头正圆柱的展开图,图 2 为正圆锥的展开图.

平面立体展开图(development of plane-surface solid)　展开图的一种. 平面立体的表面为平面,展

开图由若干个多边形(三角形、四边形等)组成. 图示



为一截头四棱柱的展开图.

不可展曲面展开图(non-developable curved surface development) 展开图的一种. 曲面上的两条连续素线是交叉的,或母线是曲线时,该曲面称为不可展曲面,如圆球、螺旋面等. 图示为圆球的近似展开.

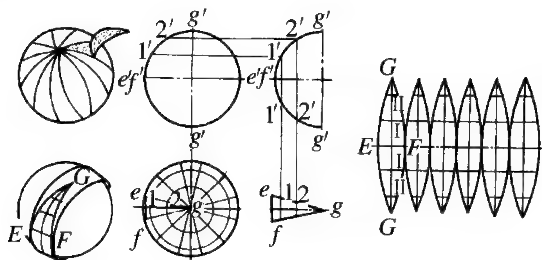
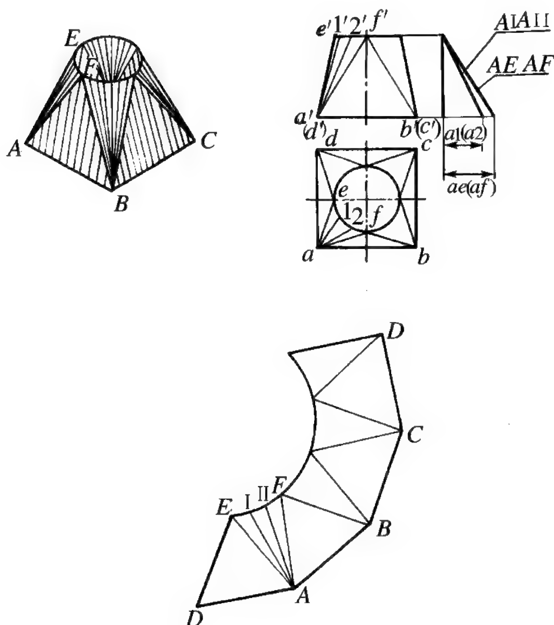


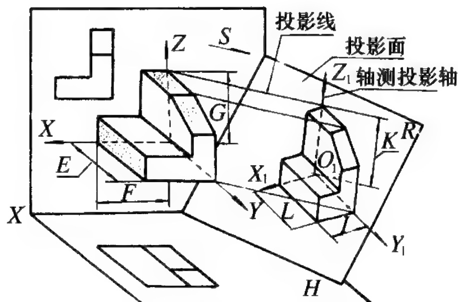
图 1　　　　图 2　　　　图 3

变形接头展开图(transition piece development) 展开图的一种. 在管路连结中,常用变形接头来连结断面不同的管子,为了使流体顺畅通过,接头内表面应力求光滑. 图示为上圆下方的变形接头,它是由四个部分锥面和四个三角形组成.



轴测投影图

轴测投影图(axonometric projection graph) 一类投影图. 用平行投影法中的正投影法或斜投影法,把空间形体连同确定该形体的空间直角坐标系一起投影到一个投影面上,这样得到的图称为轴测



投影图,如图所示. 在轴测投影图的形成过程中,要将坐标轴对投影面放成一定的角度,使在投影图上能同时反映出形体的长、宽、高三个方向. 这种图的优点是直观性较好,缺点是度量性较差,作图较繁,它一般与正投影图配合使用,以弥补正投影图直观性较差的不足.

轴测投影面(axonometric projection plane) 一类投影面. 获得轴测投影图的投影面 R 称为轴测投影面. 见"轴测投影图"条目中的图.

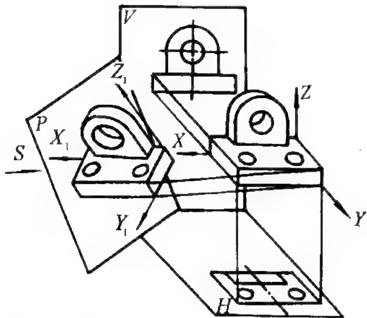
轴测投影轴(axonometric projection axes) 一类投影轴. 简称轴测轴. 见"轴测投影图"条目中的图. 坐标轴 OX, OY, OZ 在轴测投影面 R 上的投影 O_1X_1, O_1Y_1, O_1Z_1 即为轴测投影轴.

轴间角(angles between axonometric axes) 一类特殊角. 见"轴测投影图"条目中的图. 所谓轴间角,是指三个投影轴之间的夹角: $\angle X_1O_1Y_1$, $\angle Y_1O_1Z_1$, $\angle Z_1O_1X_1$.

轴向变形系数(axial projected scale) 轴测投影图中的基本参数. 见"轴测投影图"条目中的图. 投影轴与坐标轴上相应长度的比值. $I/F=p$ 为 O_1X_1 轴向变形系数; $L/E=q$ 为 O_1Y_1 轴向变形系数; $K/G=r$ 为 O_1Z_1 轴向变形系数. 轴向变形系数和轴间角是作轴测图的两个基本参数,随着物体与轴测投影面相对位置的不同,以及投影方向的改变,轴间角和轴向变形系数也就随之而改变,从而可得到各种不同的轴测投影图.

正轴测投影图(orthographic axonometric projection graph) 一类轴测投影图. 如图所示,投影方向 S 垂直轴测投影面 P,其上所得为正轴测投影图. 根据轴向变形系数不同,正轴测投影图又分为:正等测轴测投影图——三个轴向变形系数都相等($p=q=r$);正二测轴测投影图——三个轴向变形系

数中有两个相等；正三测轴测投影图——三个轴向变形系数各不相等($p\neq q\neq r$). 工程上常采用的是正等测轴测投影图和正二测轴测投影图这两种形式的轴测图.



正等测轴测投影图(isometric projection) 一种常用的正轴测投影图. 如图 1 所示，正等测轴测投影图的轴间角：$\angle X_1O_1Y_1=\angle Y_1O_1Z_1=\angle Z_1O_1X_1=120°$. 正等测轴测投投图的轴向变形系数：$p=q=r$

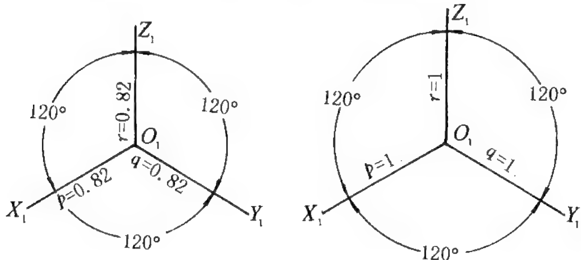


图 1　　　　图 2

$=0.82$. 在实际作图时多采用简化变形系数 $p=q=r=1$ 来代替轴向变形系数 $p=q=r=0.82$，如图 2 所示. 此时所画的轴测图略有放大，各轴向的放大率为

$$k=\frac{1}{0.82}=1.22,$$

图 3

如图 3 所示.

圆的正等测轴测投影图(circle in isometric projection) 正等测轴测投影图中常见的几何图形. 图 1 表示一立方体及其各表面上内切圆的正投

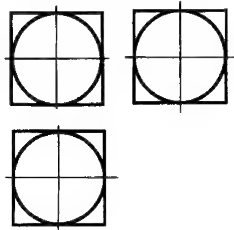
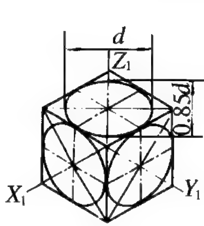


图 1

影图，这三个圆分别平行三个坐标平面且直径相等. 图 2 画出了立方体和这三个圆的正等测轴测投影图，这三个圆的轴测投影各为椭圆. 三个投影椭圆相等，投影椭圆的长轴等于圆的直径 d，短轴等于 $0.58d$，长轴和短轴分别垂直和平行于对应的轴测轴. 例如，平行于坐标平面 XOY 的圆，其轴测投影图的椭圆长轴垂直于 O_1Z_1 轴，短轴平行于 O_1Z_1 轴. 图 3 采用简化变形系数画出的同一立方体及其各表



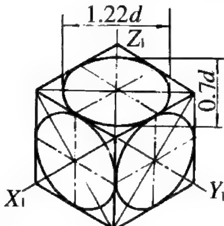
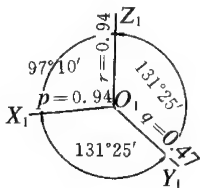


图 2　　　　图 3

面上内切圆的正等测轴测投影图，此时，椭圆的长轴等于 $1.22d$，短轴等于 $0.7d$. 圆形位于坐标面上和位于平行该坐标面的平面上，在轴测投影面上所得是相同的投影椭圆.

正二测轴测投影图(dimetric projection) 一种正轴测投影图. 如图 1 所示，正二测轴测投影图的轴



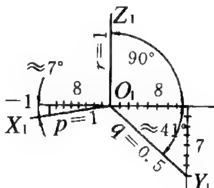


图 1　　　　图 2

间角：$\angle X_1O_1Y_1=\angle Y_1O_1Z_1=131°25'$；$X_1O_1Z_1=97°10'$. 正二测轴测投影图的轴向变形系数：$p=r=0.94, q=0.47$. 在实际作图时，多采用简化变形系数 $p=r=1, q=0.5$ 来代替轴向变形系数 $p=r=0.94, q=0.47$，如图 2 所示. 此时所画的图略有放大，各轴向的放大率

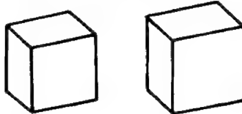


图 3

$$k=\frac{1}{0.94}=1.06,$$

如图 3 所示.

圆的正二测轴测投影图(circle in dimetric projection) 正二测轴测投影中常见的几何图形. 图 1 表示一立方体及其各表面上内切圆的正投影图，这三个圆分别平行三个坐标平面且直径相等. 图 2 画出立方体和这三个圆的正二测轴测投影图，圆的正二测轴测投影图中，两个是相同的椭圆，其长轴等于 d，短轴等于 $0.33d$. 另一个椭圆，其长轴等于 d，短轴等于 $0.88d$，长轴和短轴分别垂直和平行于对应

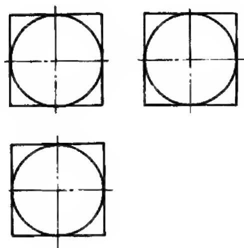


图 1

的轴测轴. 例如平行于 XOY 坐标面的圆,轴测图椭圆的长轴垂直于轴 O_1Z_1,短轴平行于轴 O_1Z_1. 图 3 是采用简化变形系数画出的同一立方体及其各表面上内切圆的正二测轴测投影图. 此时,相同的两个椭圆的长轴等于1.06d,短轴等于0.35d. 另一个椭圆的长轴等于1.06d,短轴等于0.94d. 圆形位于坐标面上和位于平行该坐标面的平面上,在轴测投影面所得是相同的投影椭圆.

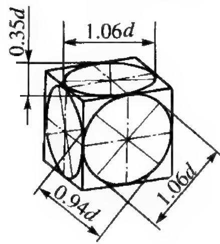


图 2

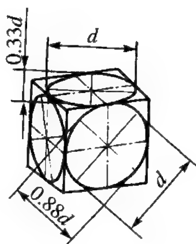


图 3

简化变形系数(simplified axial projected scale) 见"正等测轴测投影图".

斜轴测投影图(oblique projection) 一类轴测投影图. 投影方向 S 倾斜轴测投影面,所得即为斜轴测投影图. 根据轴向变形系数不同,斜轴测投影图又分为:斜等测轴测投影图——三个轴向变形系数都相等($p=q=r$);斜二测轴测投影图——三个轴向变形系数有两个相等;斜三测轴测投影图——三个轴向变形系数各不相等($p\neq q\neq r$). 工程上常采用的是斜二测轴测投影图和斜等测轴测投影图.

斜二测轴测投影图(cabinet projection) 一种斜轴测投影图. 如图所示,斜二测轴测投影图的轴间角为

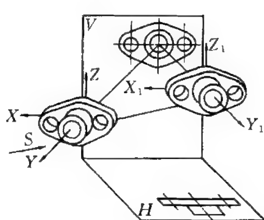


图 1

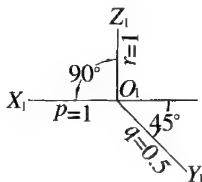


图 2

$$\angle X_1O_1Z_1=90°,$$
$$\angle X_1O_1Y_1=135°,$$
$$\angle Y_1O_1Z_1=135°.$$

斜二测轴测投影图的轴向变形系数:

$$p=r=1,\quad q=0.5.$$

圆的斜二测轴测投影图(circle in cabinet projection) 斜二测轴测投影中常见的几何图形. 图 1 表示一个立方体及其各表面上内切圆的正投影图,这三个圆分别平行于三个坐标平面并且直径相等.

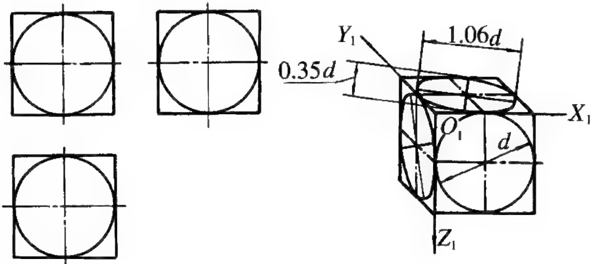


图 1　　图 2

图 2 画出立方体和这三 个圆的斜二测轴测图. 当圆所在平面平行于 XOZ 坐标平面时,其轴测图反映实形,即其投影仍是大小不变的圆. 当圆所在的平面平行于 XOY 或 YOZ 坐标平面时,圆的斜二测轴测图是椭圆,近似画法可取椭圆的长轴等于 1.06d,短轴等于 0.35d. 当圆所在平面平行于 XOY 坐标平面时,长轴与 O_1X_1 轴的夹角为 7°. 当圆所在平面平行于 YOZ 坐标平面时,则长轴与 O_1Z_1 轴的夹角为 7°. 投影椭圆的长短轴互相垂直.

斜等测轴测投影图(cavalier projection) 一种轴测投影图. 如图所示,斜等测轴测投影图的轴间角为:$\angle X_1O_1Z_1=90°$,$\angle X_1O_1Y_1=\angle Y_1O_1Z_1=135°$,斜等测轴测投影图的轴向变形系数 $p=q=r=1$.

圆的斜等测轴测投影图(circle in cavalier projection) 斜等测轴测投影中常见的几何图形. 图 1 表示一个立方体及其各表面上内切圆的正投影图,这三个

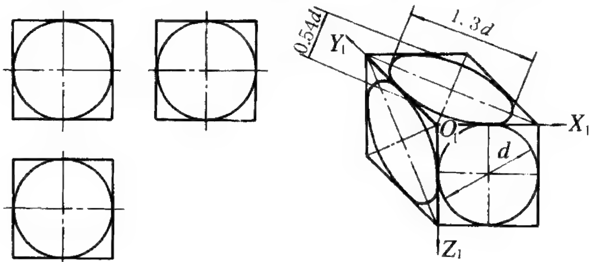


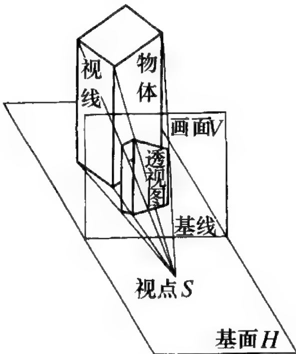
图 1　　图 2

圆分别平行于三个坐标平面且 直径相等. 图 2 画出

立方体和这三个圆的斜等测轴测图. 当圆所在平面平行于 XOZ 坐标平面时,其轴测图反映实形,即其投影是与原图大小不变的圆. 当圆所在平面平行于 XOY 或 YOZ 坐标平面时,圆的斜等测轴测投影图是椭圆,图中菱形的长对角线为椭圆长轴方向,近似画法可取长轴等于 $1.3d$,菱形短对角线为椭圆短轴方向,近似画法可取短轴等于 $0.54d$.

透视投影法

透视投影法(perspective projection method) 一种投影法. 所谓透视,即穿过透明体观察物体. 如图所示,在人与物体之间,设立一个透明的平面,称作画面(即投影面),人眼的位置 S,称视点(即投影中心),由视点至物体上各个点连线,称视线(即投影线),各视线与画面的交点,即为物体上各点的透视投影,



连结各点的投影,即得物体的透视. 由此可知,各投影线都通过视点,透视投影是一种中心投影. 根据透视投影法做出的这种立体图如同人眼所看到的物体一样,是最有真实感的描绘.

透视与一切造型艺术有关,它广泛应用于建筑设计以及其他土建工程设计,近来它也愈来愈多的应用于机械设备的造型设计. 透视的观察与研究,自古就有. 早在后石器时代所刻的甲骨上,已能找到某些简单透视现象. 在欧洲,至文艺复兴时代,透视作图方法的研究有了很大发展. 中国在透视学的研究上有显著成绩,汉代遗留的石刻与画像砖上,已有透视迹象. 宋朝李诫的《营造法式》和明朝宋应星的《天工开物》中,一些图样都应用透视原理,画出了较为成熟的透视图. 当前,立体图的作用在推广,透视投影在生产建设上将更多地显示其作用.

画面(picture plane) 一种投影面. 透视投影中,投影面常称为画面,以 V 表之. 在"透视投影法"条目中,放置物体的平面称为基面或地平面,以 H 表示. 在通常情况下,H 面为水平面,V 面与 H 面垂直,V 面与 H 面将空间分为四个分角,一般使视点 S(投影中心)位于第一分角,将物体置于第二分角.

透视的投影面(画面)可以是平面,也可以是其他面. 如圆柱面、棱柱面等,适用于表达环周的景物.

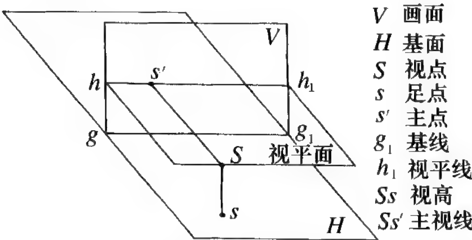
基面(ground plane) 见"画面".

基线(ground line) 亦称地平线. 透视投影术语. 指画面 V 与基面 H 的交线.

视平面(horizon plane) 透视投影术语. 通过视点的水平面称为视平面. 视平面与画面的交线称为视平线,视平线为画面上的一条水平线,其高度取决于视点的高度(参见"视点").

视平线(horizon line) 见"视平面".

视点(point of sight) 透视投影术语. 观察物体时,人眼的位置称为视点,如图中的 S. 视点在基面 H 上的投影为足点,如图中的 s. 视点在画面 V 上的投影 s' 为主点,或称为视觉中心.



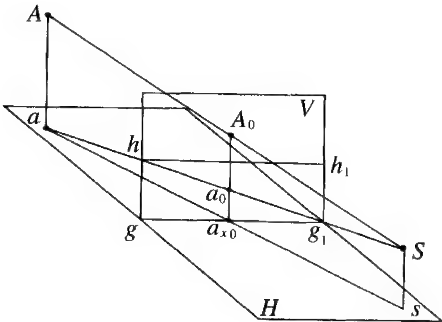
足点(station point) 见"视点".

主点(center of vision) 见"视点".

视高(height of point of sight) 透视投影术语. 视点距基面的高度称为视高,"视点"条目图中的 Ss 即是. 在一般情况下,视高等于人站在基面上,其眼睛离基面的高度(参见"视点").

主视线(axis of vision) 透视投影术语. 从视点向画面作的垂线称为主视线. 主视线的长度称为视距(参见"视点").

点的透视(perspective of a point) 一种基本几何元素的透视. 点的透视为通过该点的视线与画面的交点. 点在基面上正投影的透视称为点的次透视.



如图,A 为空间一点,a 为 A 在基面上的正投影,S 为视点,s 为足点;SA 与画面交于 A_0,A_0 即为 A 点的透视. Sa 与画面交于 a_0,a_0 即为 a 的透视,a_0 称为 A 点的次透视. 只知 A 点的透视 A_0,不能确定空间 A 点的位置,因凡是 SA_0 线上或其延长线上的点,它们的透视都可以是 A_0;如果 V,H,S 位置已确定,根据 A_0 和 a_0,就能确定 A 点在空间的位置了.

需注意图中，sa 与画面交于投影轴 gg_1 上的 a_{x0}点，A_0，a_0，a_{x0}三点在 H 面的同一条垂直线上.

点的次透视(subperspective of a point) 见"点的透视".

作点的透视(constructing the perspective of a point) 作点的透视方法. 作点的透视最常用的方

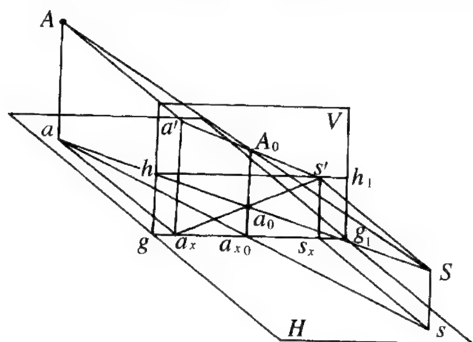
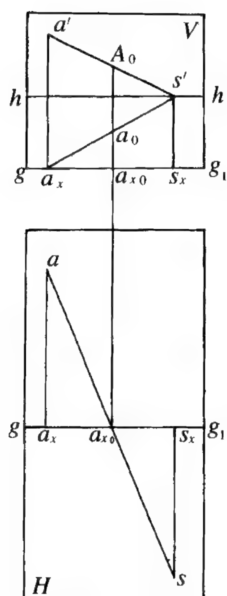


图 1



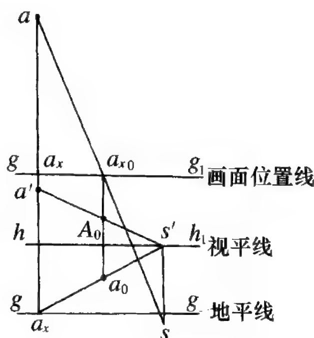


图 2　　图 3

法是利用点的正投影，求过点的视线在 V 面上的迹点. 如图 1，A 点在 H 面、V 面上的正投影分别为 a，a'；s，s'分别为足点及主点，即 S 点在 H，V 面上的正投影. sa，$s'a'$分别为视线 SA 在 H 面、V 面上的投影. sa 与 gg_1 轴交于 a_{x0}，自 a_{x0}作铅直线，与 $s'a'$交于 A_0，即为视线 SA 与 V 面的交点，亦即视线在 V 面上的迹点. A_0 为 A 点的透视. 此铅直线与 $s'a_x$ 交于 a_0，a_0 即 A 点的次透视. 若将 V，H 两面分别画出，则如图 2 所示，图中 V 面和 H 面上的 gg_1，为同一直线的两个不同位置，在 V 面上 gg_1 称地平线，在 H 面上 gg_1 称画面位置线. 作图时，常使 V，H 两面重叠，以节省所占图纸幅面，如图 3.

直线的透视(perspective of a line) 一种基本几何元素的透视. 直线的透视在一般情况下仍为直线；因为通过直线上各点的视线形成一个平面，这个平面与画面的交线，即为直线的透视. 当直线通过视点时，其透视为一点；直线位于画面上时，其透视即其本身；直线位于一个通过视点且平行于画面的平面内时，其透视在画面上无限远处.

直线的迹点(trace point of a line) 透视投影术语. 直线与画面的交点称为直线的迹点. 直线的透视必通过该直线的迹点. 当直线平行于画面时，其迹点在画面上无限远处.

直线的灭点(vanishing point of a line) 透视投影术语. 直线上无限远点的透视称为直线的灭点. 自视点发出的通向直线上无限远点的视线，与原直

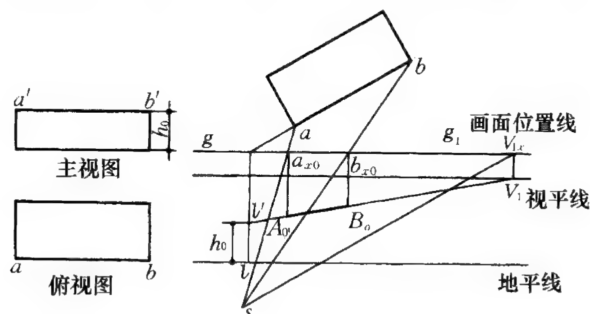


图 1　　图 2

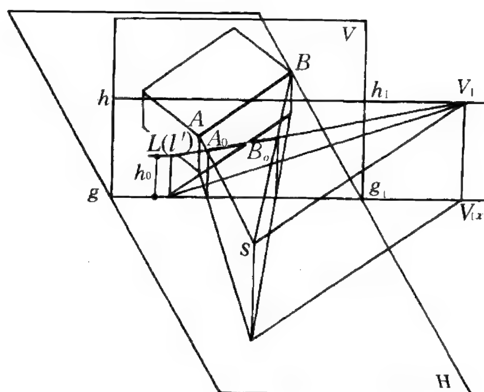


图 3

线平行，因此某已知直线的灭点就是经过视点且平行该已知直线的直线与画面的交点. 由于视平线距地平线的距离等于视点距基面的高度，所以一切水平线的灭点都在视平线上. 图 1 所示为一长方体的两视图，ab，$a'b'$为长方体上棱 AB 的两投影. 如图 2，自 s 作线平行于 ab，在地平线上得 V_{1x}，自 V_{1x}引铅直线，在视平线上得 V_1，即为直线 AB 的灭点，设以 V_1 表示. 图 3 所示为其立体图.

作直线的透视(constructing the perspective of a line) 作直线透视的方法. 作直线的透视，只要将直线上任意两点的透视连结即成. 若将直线延长，求

出其迹点和灭点，连结迹点和灭点，称为直线的全透视. 全透视系自直线与画面的交点至直线上无限远点的全部长度的透视. 如"直线的灭点"条目图 2 所示，将 ba 延长与画面交于 $L(l,l')$，L 即为直线在画面上的迹点. 连 LV 即为直线 AB 的全透视. 图中 A_0B_0 为 AB 的透视. $l'l$ 表示 AB 距基面 H 的真实高度 h_0；称为真高线.

直线的全透视(whole length of the perspective of a line)　见"作直线的透视".

真高线(true height)　见"作直线的透视".

视线法(visual-ray method)　亦称建筑师法. 作物体透视图最常用的方法.

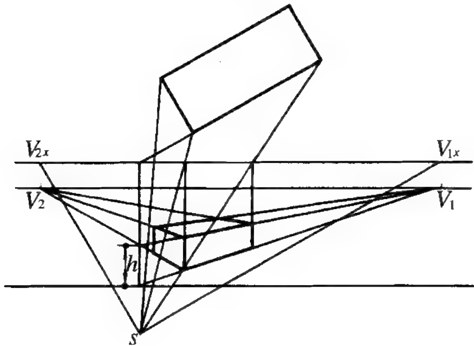


图 1

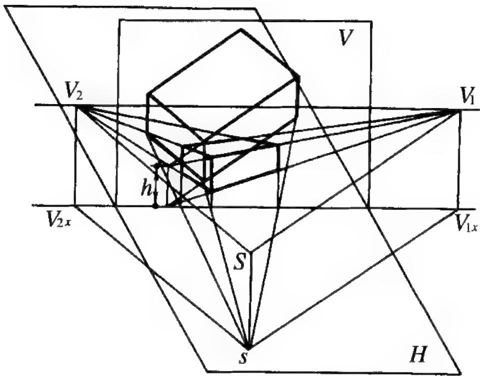


图 2

上述"作直线的透视"条目中求直线的透视所用的方法就是视线法. 又如图 1 所示，是用视线法做出的长方体的透视图. 图 2 为其立体图.

量点(measuring point)　透视投影术语. 位于视平线上的一种灭点称为量点. 利用此灭点可沿地平线量取已知水平线的真实长度，而做出水平线的透视.

如图 1，欲作基面上线段 AB(其在基面上的投影为 ab，ab 与 AB 重合)的透视：由于 A 点位于画面上，且位于基线(地平线)上，自 a 作铅直线，在地平线上即得 A 的透视 A_0；AB 为水平线，其灭点在视平线上，做出 AB 的灭点 V_1；以 a 为中心，ab 为半径作圆弧，与画面位置线交于 b_1，自 b_1 作铅直线，在地

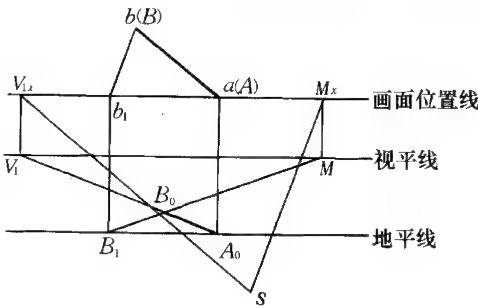


图 1

平线上得 B_1，则 $A_0B_1=ab_1=ab=AB$ 实长；作直线 b_1b 的灭点，作法为：自 s 作直线平行 b_1b，与画面位置线相交，自其交点 M_x 引铅直线，在视平线上得 M，M 即为 b_1b 的灭点；AB 的全透视为 A_0V_1，b_1b 的全透视为 B_1M，A_0V_1 与 B_1M 相交于 B_0，B_0 当为 B 点的透视；A_0B_0 即为 AB 的透视. 在图 1 中，A_0B_1 为 A_0B_0 的实长. 作 AB 的透视图时，也可先求出 M 点，其求法为：求出 AB 的灭点 V_1 后，以 V_{1x}为中心，以 $V_{1x}s$ 为半径作弧，与画面位置线交于 M_x，由 M_x 作铅直线，在视平线上得 M，然后利用 M 点及直线的实长，做出 AB 的透视图. 此 M 点称为直线 AB 的量点. 利用量点，可度量透视图中水平线的长度或

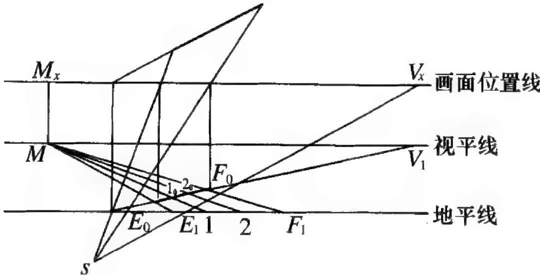
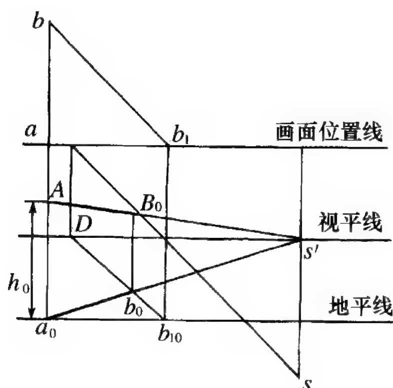


图 2

进行等分等. 如图 2，在透视图中欲将 E_0F_0 三等分，可自量点 M 引线通过 E_0，F_0，在地平线上得 E_1，F_1，E_1F_1 即 E_0F_0 的实长. 将 E_1F_1 三等分，把各分点与 M 相连，即得 E_0F_0 的三等分点 1_0，2_0.

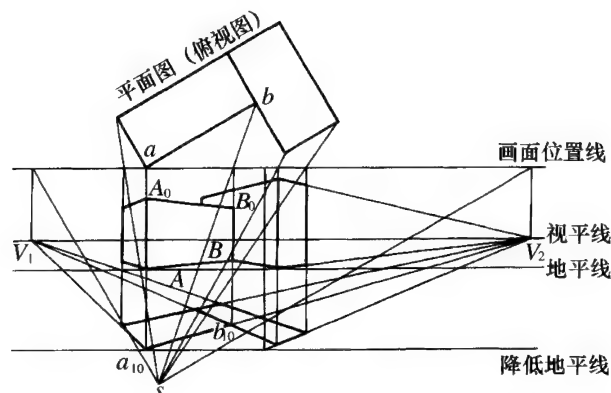
量点法(measuring point method)　作物体透视的常用方法. 在"量点"条目图 1 中求直线的透视，所用的方法就是量点法(参见"量点").

距点(measuring point for picture plane perpendiculars)　透视投影术语. 指画面垂直线的量点. 此量点到主点 s' 的距离等于视距 Ss'，习惯上将此量点称为距点. 如图，ab 为画面垂直线，且位于基面上，自 b 引 45°斜线，在画面位置线上得 b_1，自 b_1 引铅直线，在地平线上得 b_{10}，则 $a_0b_{10}=ab_1=ab$，找到 b_1b 的灭点 D，$b_{10}D$ 为 b_1b 的全透视，a_0s' 与 $b_{10}D$ 相交于 b_0，a_0b_0 即为直线 ab 的透视；a_0b_0 的实长等于 $a_0b_{10}=ab_1=ab$；D 称为画面垂直线的距点. 图中，设 AB 为距基面高 h 的画面垂直线，其在基面的投



影为 ab，则依上述求得 a_0b_0 后，由 a_0，b_0 作铅直线，利用真高 h，在 A_0s' 上得 A_0，B_0，A_0B_0 即为直线 AB 的透视，a_0b_{10} 为 A_0B_0 的实长.

降低平面图(lowered plane)　作透视图的一种方法. 当物体置于基面上时，常因视平线与地平线相距很近，图形不易作准确，且物体的次透视与物体的透视重叠混淆，因此，常采用一个比 H 面低的新基



面 H_1，利用 H_1 做出物体的透视图，称为降低平面图. 设直线 AB 在 H 上的投影为 ab，在 H_1 上的投影为 a_1b_1，a_1b_1 与 ab 所不同的仅所在空间降低. $a_1b_1 // ab // AB$，a_1b_1，ab，AB 有同一灭点及同一量点. a_1，b_1 的透视 a_{10}，b_{10} 分别与 A，B 的透视 A_0，B_0 在同一条铅直线上. 图示为利用降低平面图做出的物体透视图.

一般位置直线的透视(perspective of an oblique line)　作一般位置直线透视的方法. 求一般位置

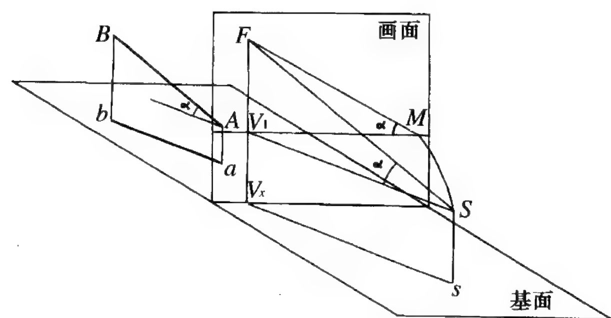


图 1

直线的透视，可求出其上两个点的透视，连结之，即得直线的透视. 如果物体上有一组平行的一般位置直线，可求出这组平行直线的灭点，以使作图简捷.

一般位置直线灭点的求法是：如图 1，设有一般位置直线 AB，其在基面上的投影为 ab，ab 的灭点为 V_1，AB 与水平面成 α 角. 自 S 作平行 AB 的视线 SF，SF 当与画面交于过 V_1 的铅直线上，交点 F 即

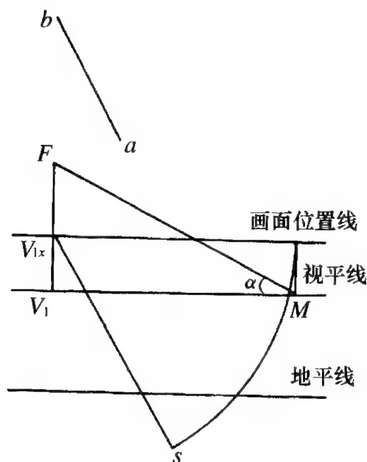
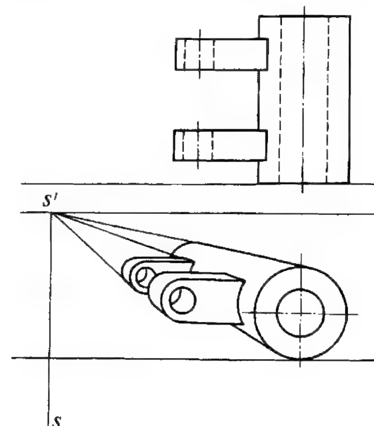


图 2

为 AB 的灭点. 以 FV_1 为轴，将$\triangle SFV_1$ 转至重合于画面的位置，此时 S 将与 ab 的量点 M 重合. 因此，欲求一般位置直线 AB 的灭点，其步骤为：如图，先求出 AB 在基面上的投影 ab 的灭点 V_1 及其量点 M，自 M 作直线与水平成 α 角，此线与过 V_1 的铅直线交于 F，F 即所求灭点.

平行透视(parallel perspective)　亦称正面透视. 一种常用的透视图. 物体的正面与画面平行的透视称为平行透视. 这种透视，多用于室内陈设、街道、



桥梁、门厅、会场等. 平行透视主要应用一个灭点，即主点，所以又称一点透视. 如图所示就是一个机器零件的平行透视.

平面的灭线(vanishing line of a plane)　一种特殊直线. 即平面上所有直线的灭点的轨迹. 经过视点且平行此已知平面的平面与画面的交线称为平面

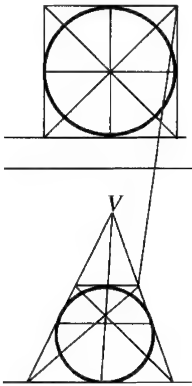
的灭线．平面上所有直线的灭点，均在该平面的灭线上．若求出平面上两相交直线的灭点，连结即为平面的灭线．

平面立体的透视(perspective of a plane solid)　作平面立体透视的方法．作平面立体的透视，就是作其表面上轮廓线的透视．平面立体的轮廓线为各种不同方向、长度、位置的直线，只要做出这些直线的透视，即可完成立体的透视．

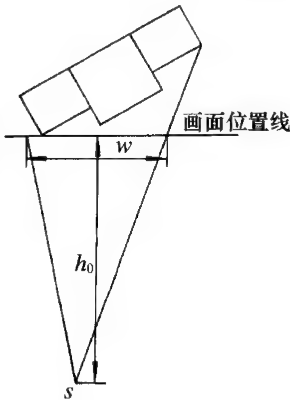
正面透视(frontal perspective)　即"平行透视"．

成角透视(angular perspective)　一种常用的透视图．物体的一个主要铅直面与画面倾斜某一角度的透视称为成角透视，一般偏转的角度为20°至30°，多用于房屋及一般物体，因为是从偏斜方向去观察物体，往往生动活泼．成角透视主要应用两个灭点，所以又称两点透视．"降低平面图"的插图即为成角透视．

圆的透视(perspective of a circle)　作圆的透视的一种方法．作圆的透视，可用外切正方形的方法．先做出圆的外切正方形的透视，再观察圆上各点在外切正方形中的位置，定出各点的透视，并连结之．一般用八点法，如图所示．由图可以看出圆心的透视不再是透视椭圆的中心．

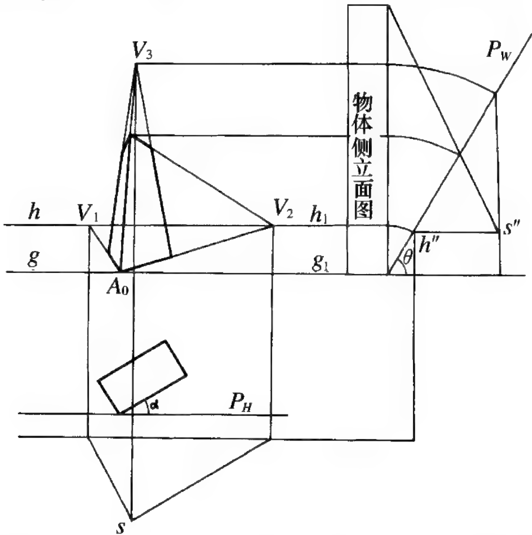


视点位置选择(location of point of sight)　选择视点位置的方法．为使透视图能将物体表现得自然生动，要妥善选定视点的位置．人眼的总视域很大，但清晰的视域仅为总视域的一部分，呈一椭圆锥形．作透视图时，视点位置的选定应使物体全部落在一个以视点S为顶点，底圆位于画面上的圆锥所围的空间内，如图，这个圆锥的底圆宽度w与其高度h_0之比可为1∶1至1∶3，此时圆锥的顶角θ约为55°至20°，而最好的方案大多是$w/h=1:2$，此时视锥的顶角θ约为30°．

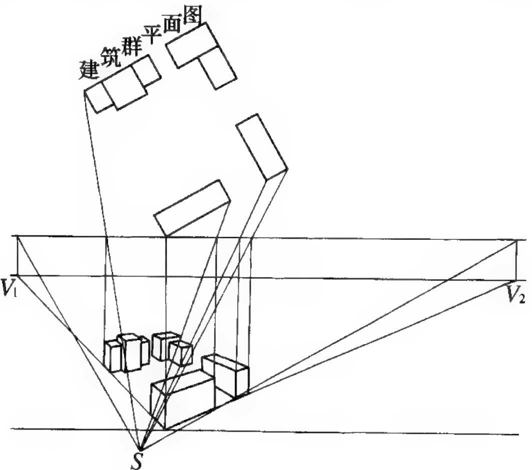


曲线的透视(perspective of a curve)　作曲线透视的一种方法．曲线的透视一般仍为曲线．作曲线的透视时，可在曲线上选取一系列能够确定和显示曲线形状的点，求出这些点的透视，用曲线光滑地连结起来即是．作平面曲线的透视或空间曲线的次透视时，通常用网格法，即先对曲线做出网格，作网格的透视，然后观察曲线上各点在网格中的位置，做出曲线上各点的透视，即可连成平面曲线的透视或空间曲线的次透视．格子的大小、稀密要看曲线的形状及曲线各处变化情况而定．

斜透视(oblique perspective)　一种透视图．斜透视是画面对物体上三个方向的主要面都倾斜的透视，多用于高层建筑物等，因其高度很大，在高度方向上也有远小近大的透视现象．斜透视应用三个灭点，所以又称三点透视．如图所示就是斜透视．



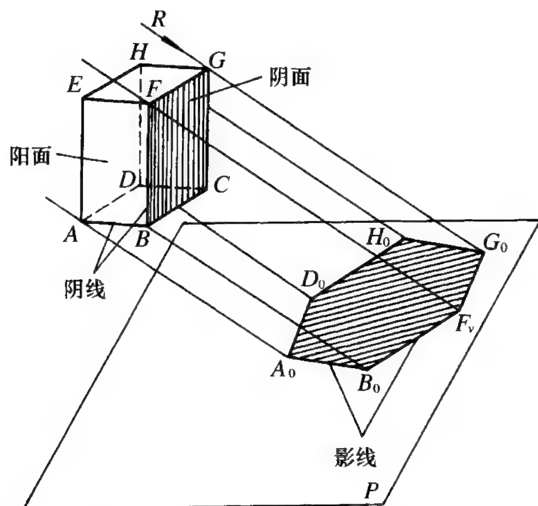
鸟瞰图(bird eye view)　一类透视图．对于范围较大的建筑群，要使视点升高，像是在空中或是在高地观看一样，这样画出的透视图称鸟瞰图，如图．



阴　影　法

阴影法(method of shade and shadow)　一种投影法．物体在光线照射下产生的暗区称为阴(shade)，而在其他物体表面(例如地面)产生的影子

称为影(shadow).人们在图形上画出阴影,可以明显地反映出物体各部分的凹凸、深浅、明暗,使图面生动,富有立体感,丰富所画图形的表现能力.这种利用图形阴影的方法称为阴影法.在正投影多面视图、轴测图、透视图上都可以添绘阴影.



图中表示阴影的各部分名称.图中 R 表示光线方向,物体 $ABFE$,$EFGH$,$ADHE$ 三个面受到光的照耀称为阳面;$ABCD$,$BCGF$,$CGHD$ 三个面光线照射不到称为阴面.折线 $ABFGHDA$ 是阴阳面的分界线称为阴线.物体的影子落在 P 面上,P 面称为承影面.影子的轮廓线称为影线.由图可以看出,影线是阴线的影子,影是由阴线的影子所围成.

阴线(shade line)　见"阴影".

影线(shadow line)　见"阴影".

习用光线(conventional ray of light)　阴影法术语.指正投影多面视图绘影时规定的一种平行光线.在物体的正投影多面视图上添绘阴影时,为便于作图和度量,通用一种特别方向的平行光线称为习用光线.如图 1,设立方体的各棱面分别平行于相应的投影面,则立方体的对角线 AO 表示习用光线的方向,习用光线的方向是从立方体的左上前角射向右下后角,AO 对各投影面的实际倾角都相等,等于 35.264°,习用光线的三面正投影均与投影轴成 45°,如图 2.绘制阴影时,物体要一律置于第一分角.

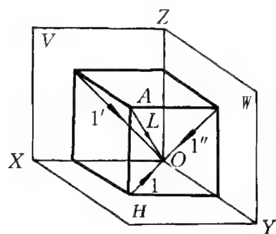


图 1

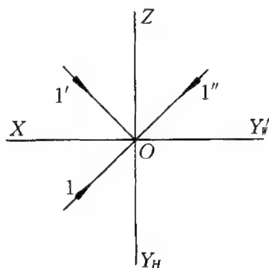


图 2

正投影多面视图中基本几何元素的影子(shadow of point, line and plane in orthographic multiview)　阴影法术语.作点的影子即求过点的习用光线与承影面的交点.作直线的影子即求过直线的光平面与承影面的交线.作曲线的影子连结曲线上各点的影子即成.作平面的影子可求围成平面的各线的影子.上述均可应用画法几何的原理与作法完成.

正投影多面视图中基本体的阴线(shade line of basic solid in orthographic multiview)　阴线的判定法则.平面立体的阴线一般根据习用光线的方向即可判定.基本曲面立体的阴线的判定如下:

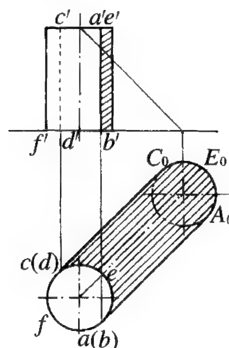


图 1

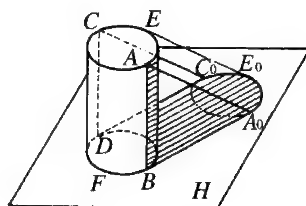


图 2

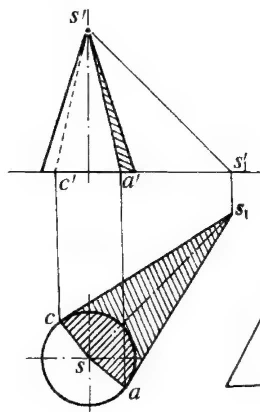


图 3

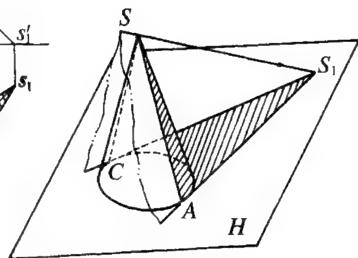


图 4

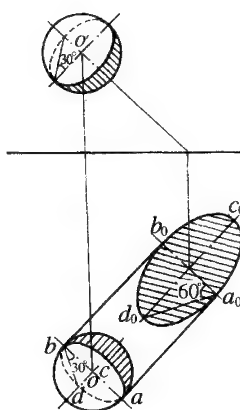


图 5

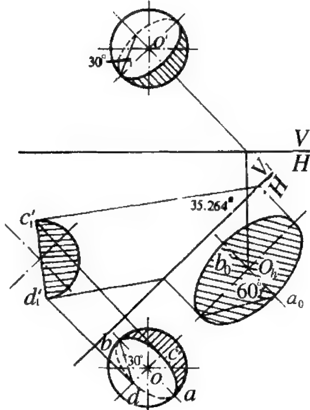


图 6

1. 如图 1，欲确定圆柱的阴线，在俯视图中作习用光线的水平投影，使其与圆柱的水平投影相切，即得阴线的水平投影 ab，cd 和正面投影 $a'b'$，$c'd'$；图 2 为其立体图. 圆柱的整个阴线为 $ABFDCEA$，包括圆柱的两条素线 AB，CD 和柱底的两个半圆 AEC，BFD.

2. 如图 3，欲确定圆锥的阴线，过锥顶 S 作习用光线 $SS_1(ss_1, s's'_1)$，SS_1 与 H 面的交点 S_1，即为锥顶 S 落在 H 面上的影子. 自 s_1 作直线与圆锥底圆相切，得切点 a，c，由 a，c 得 a'，c'，则 $SA(sa, s'a')$，$SC(sc, s'c')$ 为圆锥侧表面的阴线. 图 4 为其立体图.

3. 如图 5，圆球的阴线为垂直习用光线方向的球上一个大圆，此大圆的水平投影和正面投影是椭圆，在水平投影中，投影椭圆的长轴为 ab，长度等于圆的直径，短轴与长轴垂直. 简单的作法是：自长轴端点 a，b 作 30°角，即可定出短轴 cd. 同法，可做出阴线的正面投影椭圆. 图 6 为应用变更投影面法，设立平行于习用光线方向的新 V_1 面，在 V_1 面上，光线反映真实倾角 35.264°，此时球面上的阴线大圆投影成直线 $c'_1d'_1$.

正投影多面视图中的阴影（shade and shadow in orthographic multiview） 阴影法术语. 指物体多面视图添绘阴影的方法. 在正投影多面视图中加绘物体的阴影，实际上是画出阴和影的正投影. 图 1 是窗口的两视图，图 2 是画出阴影的窗口的两视图. 图 3 为台阶的阴影；图 4 为带方盖的圆柱的阴影.

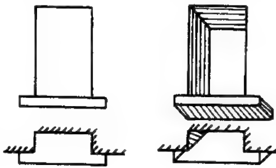


图 1　　图 2

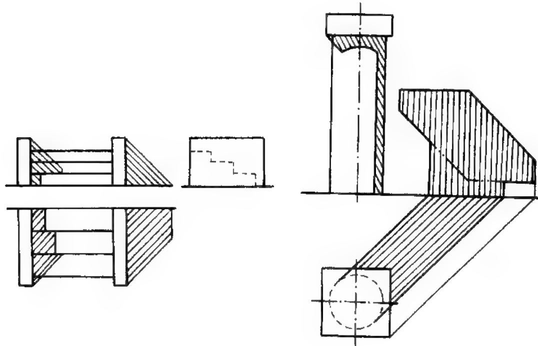


图 3　　图 4

在物体的正投影多面视图上添绘阴影，明显的可以增强立体感和艺术性，但作起来费时费力. 先要区别其上的阴阳面，定出阴线；然后求阴线各段落在不同承影面上的影子，逐一进行，才能完成物体的阴影，所以归纳阴线求法的规律和不同线段落在不同表面上的影子的规律，得出简捷作法. 这是一项需要探索的项目，过去已经有了一些，例如圆球的阴影便是分析归纳出的简捷作法，但还需继续进行研究.

轴测图中为物体添绘阴影（determine shade and shadow of solid object in axonometric drawing） 阴影法术语. 指为物体轴测图添绘阴影的方法. 轴测图中添绘阴影，光线的方向可根据物体的具体情况任意择定. 图 1 是台阶的阴影，图 2 是带方盖的圆柱的阴影；光线方向分别示于图的右上角.

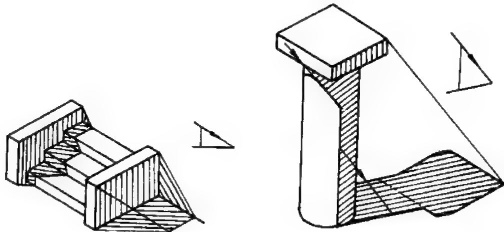
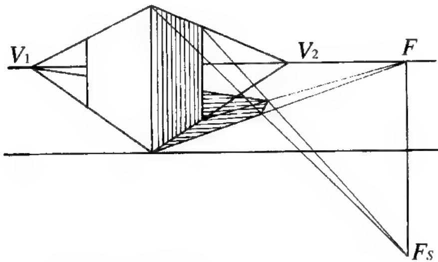


图 1　　图 2

透视图中为物体添绘阴影（determine shade and shadow of solid object in perspective drawing） 阴影法术语. 指为物体透视图添绘阴影的方法. 透视图中绘阴影，光线方向可根据物体具体情况任意选定. 光源可设在无限远处（如太阳光线），也可设在有限距离处（如灯光）. 在正投影多面视图中添绘阴影的规律与方法，仍可应用，但应结合透视的消失规律. 图中，V_1，V_2 为物体透视的两个灭点，F_S 为光线（平行光线）的灭点，F 为光线次透视的灭点.



工程制图

工程制图（engineering drawing） 研究工程图样的学科. 按一定的投影方法和技术规定将已有的或构想的机器、设备、建筑物等的形状、大小和有关资料表达在纸面上，称此纸样为工程图样.

图样是表达、交流思想的工具，任何工程的进行都离不了图样. 比如，在设计改进机器设备时，要通过图样表达设计思想和要求，在制造机器过程中，无论是制坯、加工、检验、装配等各个环节，都要以图样作依据，使用机器时，也要使用图样来帮助了解机器的结构和性能；在房屋建造中，从研究设计到工程竣工，都要广泛地应用图样. 因此图样被认为是工程技

术上的一种语言.而画法几何所阐述的理论则是这“工程语言”的语法.

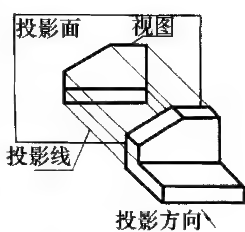
在人类进化中,语言的产生为人类思想的相互沟通奠定了基础,文字的出现,又补充了语言表达的不足,而对于具体物体形状的描述,单靠语言文字往往不能使人理解得清楚,用图样就能表达得准确和完善.图样早在远古时代就在各地人类的生产活动中应用.其中正式用于工程且留传至今的,最著名的有宋代李诫所著《营造法式》(公元 1100 年),书中附图精湛,已有各种图示方法;再如,明宋应星的《天工开物》(公元 1637 年),书中用到与现代所用颇为相似的各种图示方法.近几十年,随着生产的发展、科学技术的进步,制图科技在迅速发展.其间中国有关学者与科技人员作了大量工作,并有了许多创新成果,适应了各项建设以及大型工程的需要.当前,计算机绘图在广泛应用和发展;交互图学已遍及许多国家.

工程图学(engineering graphics) 研究工程图形理论和方法的学科.内涵广泛,包括除去纯艺术图以外的所有各种制图.包括从一个机件的多面正投影、各种图解方法、以及图解微积分等;包括以几何理论为基础的投影制图和以数学推导为基础的非投影制图;它所包括的计算机绘图与计算机辅助设计密切关连.

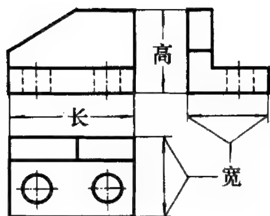
机械图(machine drawing) 一种常见的工程图.机械图是机器设备设计、制造与生产各个环节中使用图样的通称.机械图主要有两种:一是零件图;二是装配图.表达一个零件的图样称为零件图,表达整个机器或部件的图样称为装配图.

建筑图(architectural drawing) 一种常见的工程图.它是建筑工程使用的各种图样的通称.建筑图按设计阶段,主要可分初步设计图、设计图和施工图;为了表示建筑物的外部和内部构造,有平面图、立面图、剖视图、大样图等.

视图(view) 工程制图的基本概念之一.指机件向投影面投影所得的图形.如图所示.设有一投影面,将机件置于投影面与观察者之间,按正投影条件(即相互平行的投影线垂直于投影面),把机件上可见的轮廓线画成粗实线,把不可见的轮廓线画成虚线,如此得到的投影图就称为视图.



三视图(three-view drawing) 一种常见的视图.基本视图中主视图、俯视图和左视图的总称.基本视图虽有六个,但并非每一个机件都要画六个视图,而是根据机件形状的复杂程度选定,六视图中最基本、又最常用的是主视图、俯视图和左视图,这三个视图既反映了机件长、宽、高三个方向的形状,又是三个主要的投影方向,故一般用三视图来表达机件的形状.画三视图时,应使几个视图间的总的长、宽、高和每一部分的长、宽、高都保持长对正、高平齐、宽相等的投影规律,如图.



基本视图(basic view) 一类常见的视图.机件

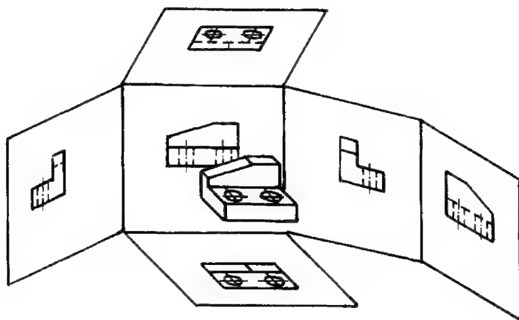


图 1

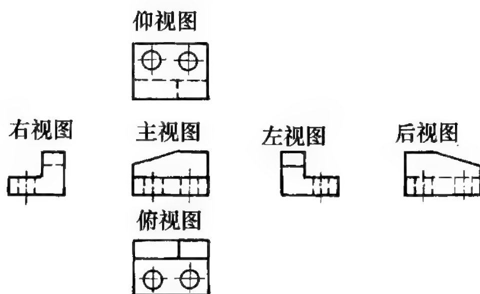


图 2

向基本投影面投影所得的视图称为基本视图.国家标准 GB 中把正六面体的六个面规定为基本投影面,机件放置在正六面体的中间,分别从机件的前、后、左、右、上、下六个方位向六个基本投影面投影,再把各投影面如图 1 展开,就得到六个基本视图.各基本视图的位置关系如图 2.

主视图(front view) 一种视图.指从机件的前方向后投影所得的视图.

俯视图(top view) 一种视图.指从机件的上方向下投影所得的视图.

左视图(left side view) 一种视图.指从机件的左方向右投影所得的视图.

右视图(right side view) 一种视图.指从机件的右方向左投影所得的视图.

仰视图(bottom view) 一种视图.指从机件的下方向上投影所得的视图.

后视图(rear view) 一种视图.指从机件的后

方向前投影所得的视图.

斜视图(oblique view)　一种常见的视图.当机件上某部分的结构形状与基本投影面处于倾斜位置时,将机件的倾斜部分向不平行于任何基本投影面的平面投影所得的视图称为斜视图.画斜视图时,必须在斜视图上方水平标出它的名称"×向",在相应

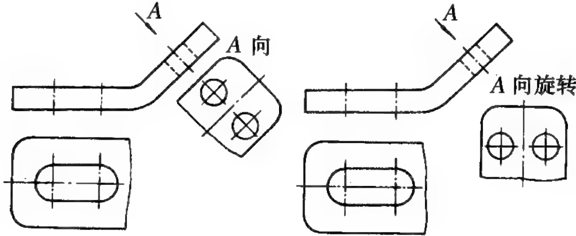
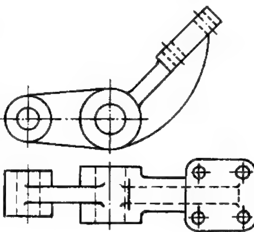


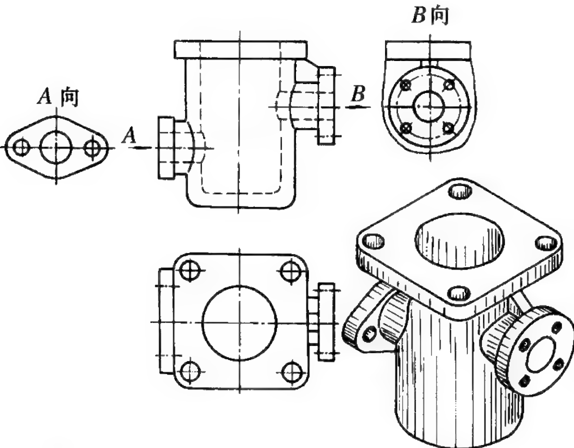
图 1　　图 2

视图附近用箭头指明投影方向,并注上同样的字母.斜视图一般按投影关系配置,如图 1,必要时也可画在其他位置.为了画图方便,也允许将图形旋转,标注形式为"×向旋转",如图 2.斜视图只要求画出倾斜部分形状,其余部分不必画出,可用波浪线断开.

旋转视图(aligned view)　一种常见的视图.当机件的倾斜部分有一旋转轴线时,假想将机件的倾斜部分旋转到与某一选定的适宜投影面平行后,再向该投影面投影所得的视图称为旋转视图.画旋转视图时一般不标注.如图.



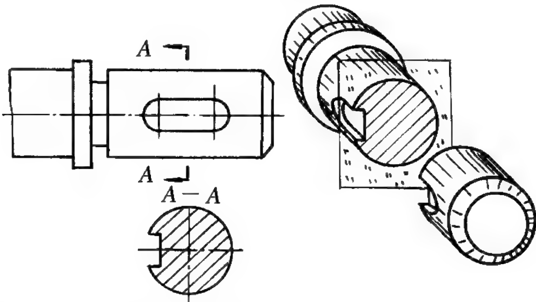
局部视图(partial view)　一种常见的视图.当机件已采用一定数量的基本视图表达后,仍有部分形状未表示清楚,这时,将机件的这一部分单独向基



本投影面投影所得的视图称为局部视图.

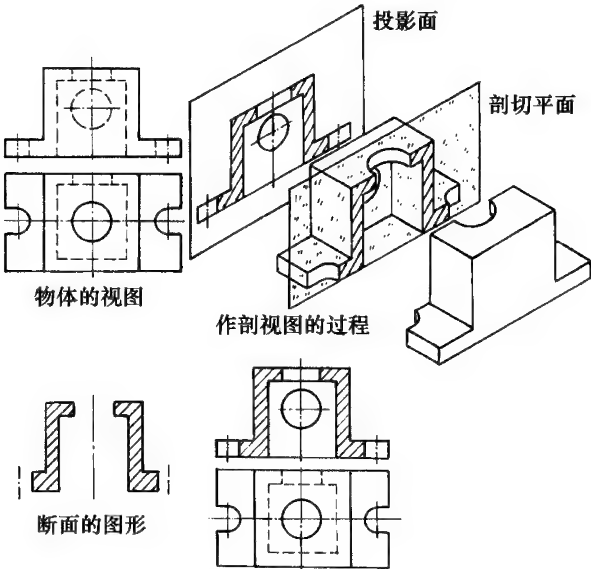
画局部视图时,一般在局部视图上方水平标出视图的名称"×向",在相应的视图附近用箭头指明投影方向,并注上同样的字母,如图中的 *B* 向.但局部视图按投影关系配置,中间无其他图形隔开时可省略标注.局部视图的断裂边界用波浪线表示.当所表示的局部结构是完整的,且外轮廓线又成封闭时,波浪线可省略不画,如图中的 *A* 向.

剖面图(section)　一种常见的视图.假想用剖



切平面将机件的某处切断,仅画出断面的图形称为剖面图.如图所示,画剖面图时,为了反映断面的真实形状,剖切平面一般应垂直于轮廓线或轴线剖切.

剖视图(sectional view)　一种常见的视图.假想用剖切面(一般用平面)剖开机件,将处在观察者和剖切面之间的部分移去,而将其余部分向投影面



投影所得的图形称为剖视图.国家标准中规定,剖切面与机件接触的断面图形要画出剖面符号,图中所画 45°细实线部分就是一种剖面符号.不同的材料采用不同的剖面符号.机件的不可见形状虽然可用虚线表示,但当机件的内形比较复杂时,虚线较多,会使图形很不清晰,不便于画图、看图和标注尺寸.如果假想用剖切平面剖开机件,则它的内部形状和层次关系就能清晰地显示出来.

画剖视图时,一般应在剖视图上方水平标出剖视图的名称"×—×",在相应视图上用剖切符号(线宽 1～1.5*b* 的粗实线)在起、迄和转折处表示剖切位置,用箭头表示投影方向,并注上同样的字母.当剖

视图与相应视图具有投影关系，中间又没有其他图形隔开时，可省略箭头. 当单一剖切平面通过机件的对称平面或基本对称的平面，且剖视图按投影关系配置，中间又没有其他图形隔开时，可省略标注，如图.

零件图(detail drawing) 一种常见的工程图. 即表达零件外形和内部构造的图样，在零件制造过程中用于加工和检验等工作. 零件图反映零件制造过程中的全部技术资料，包括视图、尺寸、技术要求及标题栏等.

装配图(assembly drawing) 一种常见的工程图. 指表达一部机器或部件的图样. 是设计、装配、检验、使用和维修机器的重要技术资料. 其中有表达机器(或部件)的工作原理、装配关系及主要零件的结构形状的一组视图，装配、检验、安装时所需的尺寸和技术要求，以及标题栏和明细表.

计算机绘图

计算机绘图(computer graphics) 一种绘图方法. 是利用计算机进行数据处理、提供图形信息、控制绘图设备、实现自动绘图. 又称为数控绘图或自动绘图，简称 CG. 是 20 世纪 60 年代初迅速发展起来的一门新兴学科. 计算机绘图具有高速度、高精度的特点，不但可以代替繁重的手工绘图，而且可以绘制远非手工能胜任的高精度的复杂图形.

计算机绘图包括静态绘图和动态绘图两种形式. 前者绘图过程操作者不能干预，若要变更图形，必须修改程序重新绘制，因而称为被动式绘图；后者称为交互式计算机绘图.

计算机绘图涉及多种学科的知识，例如不同工程技术的专业知识、图学知识、数学知识(如矩阵论、微分几何、计算几何等)、程序设计、计算机科学基础等.

交互式计算机绘图(interactive computer graphics) 一种动态式计算机绘图，简称 ICG. 也称为动态绘图. 绘图过程中，操作者随时可根据需要控制、干预正在显示的图形，直至满足要求为止. 动态绘图还可用于动画、模拟、仿真等技术.

当人们使用计算机绘图时，人与计算机之间可即时交换信息，而且这种交换信息是双向的. 比如，计算机在执行程序中，将会及时告知操作者进行情况，还可询问下一步作什么；如有更改，即可另给参数重新执行. ICG 也可叫做“人-机对话”.“交互式”功能对计算机绘图和计算机辅助设计非常重要，它给 CG、CAD 进行过程增加了活力.

计算机图形显示(computer graphics display) 计算机绘图术语. 是计算机所绘图样在屏幕上的显示. 计算机绘图一般指在计算机的控制下，从绘图机上输出永久性的图形. 适用于静态绘图. 如果图形输出设备不是绘图机，而是阴极射线管(CRT)构成的显示器，则称为计算机图形显示，简称 CGD. 它既适用于静态绘图，也适用于动态绘图，输出的是供临时观察或实时监视的屏幕图形.

计算机辅助设计(computer aidid design) 计算机绘图术语. 是应用计算机进行设计. 设计师以计算机为设备，完成设计、计算、生成图形，并输入其他有关资料的过程，简称 CAD. 即从产品的概念设计、生产设计，直到制造为止的整个设计过程都是以计算机为主要技术手段进行工作.

计算机绘图系统(computer graphics system) 一种计算机系统. 是实现计算机绘图必须具备的以计算机为主的全部硬件系统和软件系统的总称. 硬件系统包括中央处理机和输入、输出设备等，软件系统包括系统软件、绘图软件、数据库、图形库等.

图形输入板(graphic input tablet) 一种图形数据的采集装置. 可以将图上的点、线转换成坐标数据输入到计算机.

数字化仪(digitizer) 计算机绘图的设备. 其结构与用途类似于图形输入板，但其板面更大. 主要用作输入更复杂的图形. 可以将图形转换成坐标数据形式，存贮在计算机内存或磁盘、磁带上，也可以在显示器或绘图机上复制成图.

绘图机(plotter) 计算机绘图系统的主要设备. 在计算机的控制下，能根据计算机发出的脉冲信号，完成各种绘图动作的绘图设备. 常见的有滚筒式绘图机、平台式绘图机. 是绘图系统的主要设备.

计算机硬件设备(computer hardware device) 计算机主机及其外围设备，如键盘、屏幕显示器、打字机、绘图机等. 计算机还必须有系统软件才能工作.

计算机系统软件(computer system software) 计算机科学术语. 指计算机操作系统及进行计算的支撑软件，如 BASIC 语言、FORTRAN 语言等. 具备系统软件即可进行科学计算. 若欲绘图，还要有绘图软件.

绘图软件(plot software) 计算机绘图术语. 指用计算机语言编写的具有各种绘图功能的子程序的集合，包括基本绘图软件、通用绘图软件和应用绘图软件.

基本绘图软件(basic plot software) 一种绘图软件. 具有计算机绘图系统基本功能的软件，如抬笔移动、落笔画线、书写字符等最基本的一些子程序.

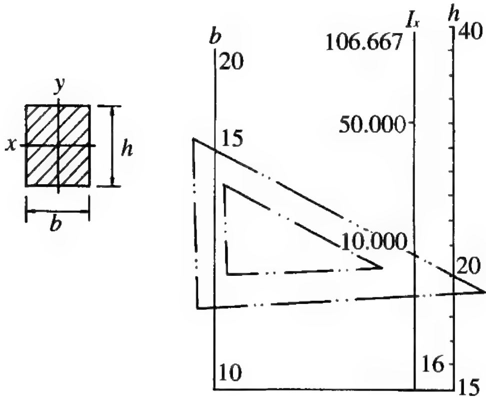
通用绘图软件(general plot software) 一种绘图软件. 在基本绘图软件的基础上，具有进一步绘

图功能的软件，是常见图形的子程序，如画多边形、画弧线、几何元素求交的子程序.

应用绘图软件(special plot software)　一类绘图软件.针对各工程技术的需要，对特定的设计与绘图问题，研制出的各专门软件.在这方面，可开拓的研究范围非常广阔.它研制出的成果，为科学技术服务，使计算机绘图向深度、广度有效和迅速地发展.

图　算　学

图算学(nomography)　应用数学的一个分支.是19世纪兴起的学科.研究根据数学原理，绘制由各变量的图尺组成的计算图，用以解决重复应用的方程式各答案的学科.计算图是根据方程式各变量的函数关系，由相互间有一定几何位置、并刻有分度标值的直线、圆或曲线图尺所组成.计算图又叫诺谟图，是由外文音译而成，法文为 Nomographe 或 Nomogramme，希腊文是由 Noμos(规律)和 Γραφω(描绘)两个字合成，英文为 Nomograph 或 Nomogram.计算图是1895年法国工艺学院教授奥根(Ocagne)所提出，其后发展成为独立学科.



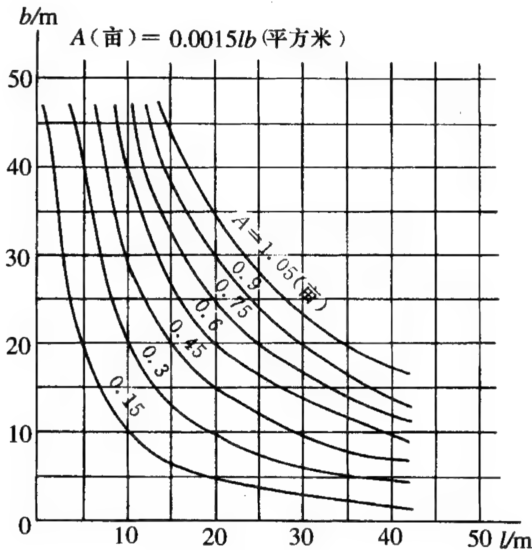
计算图广泛地应用于工程技术、自然科学各个领域.图算学是应用数学的一个分支，也是工程图学的一个内容.在中国现代化建设中，设计部门、工矿企业、科研机构都有大量计算工作，若广泛应用计算图，可以节省大量时间和人力.图算有其特点，在计算机发展的今天，实用价值依然很大.例如，在力学中常见的矩形的截面特性有轴 x 的抗弯惯性矩

$$I_x=\frac{bh^3}{12},$$

这只要绘出计算图(如图)，任意给出一个 b 值、一个 h 值，用直尺一量，立即就得出 I_x 的数值.这样，计算图是以简单的动作，代替了复杂的计算.对一般工程计算，其精确度可以满足要求.在工程设计中，经常要变更已知数值，以适合设计要求，这样，利用计算图是很方便的.例如，对于轴 x 的抗弯惯性矩，用一根直尺，就可以选择合宜的 b,h 值，得出所需要的 I_x 值.

计算图(nomogram)　见"图算学".

共点计算图(concurrent chart)　一类计算图.所谓共点计算图，是指在平面坐标系中绘制函数式的几何图形.在均等标值的坐标系中，除一次函数式



$y=mx+b$ 外，函数图形均为曲线.如图，方程式 $A=0.0015lb$，其中土地长为 l，宽为 b，图为将 A 的单位由平方米换算为亩的共点计算图.例如，由图可知土地长 $l=20$m，宽 $b=10$m，面积 $A=0.3$ 亩.为使用准确与方便，图中的曲线图尺可以校直.

函数图尺(functional scale)　计算图的组成元素.按方程式变量的函数值顺次依大小比例标在图尺上，称函数图尺.而图尺上标出的数字是变量的数值.例如，函数式为 $F(u)=u^2$，可列表计算如下：

变 量 u	0	1	2	3	4	5
变量的函数值 $F(u)=u^2$	0	1	4	9	16	25
$M_uF(u)=m_uu^2$ (设 $m_u=2$)	0	2	8	18	32	50

作图时，设以2mm为长度单位，此2mm称图尺系数.u 图尺的图尺系数以 m_u 表示.$F(u)=u^2$ 的函数图尺如图1.

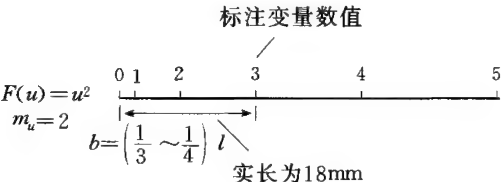


图1

图尺分均等图尺与不均等图尺，各标值点的间距相等为均等图尺，如 $F(x)=3x$，设 $m_x=3$，如图 2 为均等图尺；图 1 为不均等图尺.

x	0	1	2	3	4	5
$F(x)=3x$	0	3	6	9	12	15
$m_x f(x)=3\times 3x=9x$	0	9	18	27	36	45

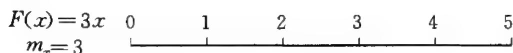


图 2

共线计算图(alignment chart)　一类计算图. 所谓共线计算图，是指计算图各图尺上符合方程式的一组变量数值位于一条直线上. 如"图算学"条目中的插图，$b=15\text{cm}$，$h=20\text{cm}$ 时，$I_x=10.000\text{cm}^4$，则此三标值即位于同一条直线上，这条直线叫"指示线"，指示线可以是直线、圆弧，也可以是两条平行线或互相垂直的两直线.

常用的共线计算图有：平行图尺计算图、Z 形图尺计算图、汇交图尺计算图、有曲线图尺的计算图、有圆形图尺的计算图、有网线图尺的计算图、比例图尺计算图、平行指示线计算图、垂直指示线计算图、圆弧指示线计算图.

指示线(index line)　见"共线计算图".

图尺系数(scale coefficient)　见"函数图尺".

均等图尺(uniform scale)　见"函数图尺".

不均等图尺(non-uniform scale)　见"函数图尺".

线性分式函数图尺细分度(scale subdivision of linear fractional function)　图算学术语. 指一种图尺的细分度方法. 若线性分式函数

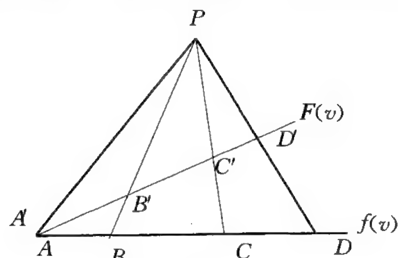


图 1

$$F(v)=\frac{af(v)+b}{cf(v)+d}$$

中，$ad\neq bc$，则 $f(v)$ 与 $F(v)$ 互成射影变换，两函数均作成直线图尺时，可根据 $f(v)$ 图尺上的标值刻度做出 $F(v)$ 图尺上的标值刻度. 如图 1，设 $f(v)$ 上 A，B，C，D 的标值 v_1，v_2，v_3，v_4 已知，$F(v)$ 上 A'，B'，C' 的标值分别为 v_1，v_2，v_3，而 D' 标值为 x，则 x 的标值应为 v_4.

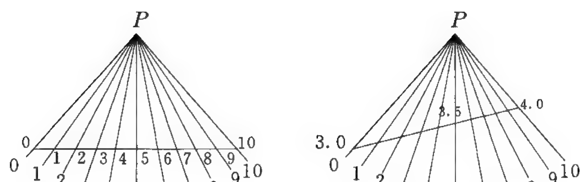
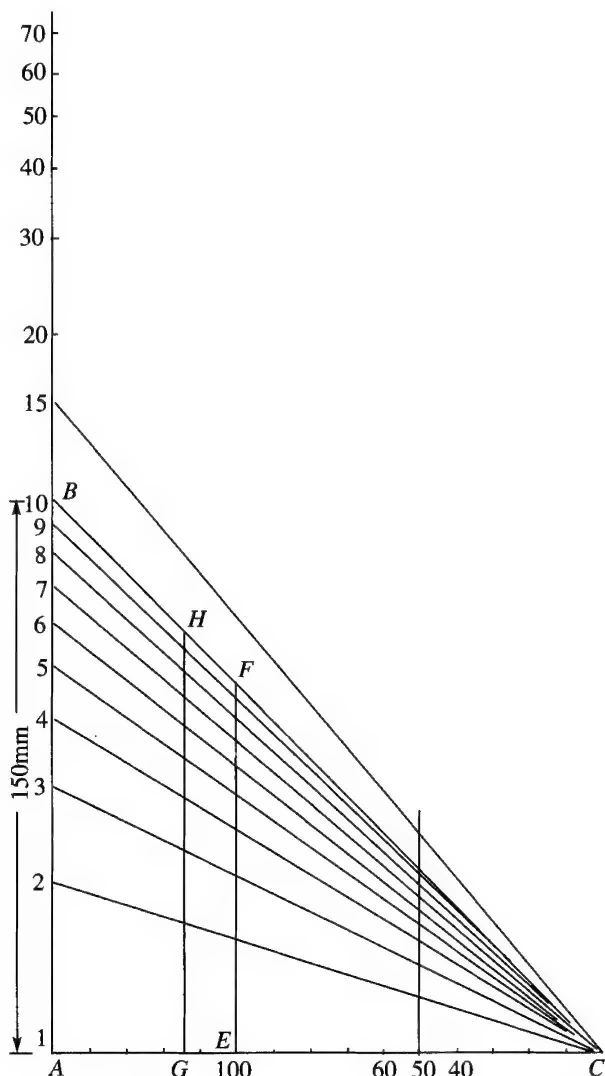


图 2　　图 3

在这样的情况下，不均等图尺的细分度可如下完成：首先在透明纸上，准备一个射影线束，如图 2，假如要将图尺 3 与 4 之间的各个 1/10 的分标值点做出，则将线束的 P-0，P-5，P-10 通过 3.0，3.5，4.0，如图 3，用针尖可得各 1/10 标值点. 此系以图尺的"交比相等"为前提. 如果这样，则细分点的位置是准确的，不然，则紧密的近似.

对数图尺三角形(triangle of logarithmic scale)　图算学术语. 即制作对数图尺的方法. 为了能够便利地绘出对数图尺，如图，设图尺系数为 150mm，画 $AB=150\text{mm}$，作 $AC\perp AB$，并使 $AC=AB$，$\triangle ABC$



即为对数图尺三角形. 在 AB 上按 lg1 到 lg10 的值，

注出 1 到 10 的标值，将各标值点与 C 相连，得各斜线. 在 AC 上标出 40，50，60 等各位置，自 40，50，60 等点作直线平行于 AB，此直线与各斜线相交，即得图尺系数为 40，50，60…时各对数图尺 1 到 10 的标值点. 图中 EF 为图尺系数为 100 时的对数图尺，GH 为图尺系数为 115 时的对数图尺. 在图中的对数图尺三角形上，若将 AB 向上延长，则可标出 10 到 100 的标值. 延长 AB，使所延长的一段等于 AB，将 AB 上的各标值点移到延长的这段长度上，即得图尺上的 20，30，40，…，90，100 各标值点. 由于所延长的这段与 AB 的分度相同，所以实际上不需将 AB 延长，而直接以 AB 上的 1 到 10 作 10 到 100 用. 又因为 AB 上的标值是在图尺系数为 150mm 时做出的，若仍用此图尺，而图尺系数改为 50mm，则原图尺的系数 150 的标值 2 应改为 8，标值 3 应改为 27 等.

平行图尺计算图(parallel scale chart)　一种常用的计算图. 方程式各变量的图尺是平行的直线，称为平行图尺计算图. 最简单的平行图尺计算图具有三条平行的直线图尺，适用于方程式$F_1(u)=F_2(v)+F_3(w)$. 在这样的计算图上，设在图尺 $F_2(v)$，$F_3(w)$上任取 v_1，w_1，连以直线 v_1w_1，在$F_1(u)$图尺上得 u_1，则$F_1(u_1)=F_2(v_1)+F_3(w_1)$. "图算学"条目中的图便是这样的平行图尺计算图，该计算图系将 $I_x=bh^3/12$ 变为

$$\lg I_x=\lg \frac{b}{12}+\lg h^3,$$

$$\lg I_x=(\lg b-\lg 12)+3\lg h$$

而做出的. 平行图尺计算图的原理和作法是：如图 1，在$F_2(v)$图尺和$F_3(w)$图尺上，取 v_0，w_0 为标值起点，连 v_0w_0，在 $F_1(u)$图尺上得 u_0，则 $v_0u_0w_0$ 为标值基线，$F_1(u_0)=F_2(v_0)+F_3(w_0)$.

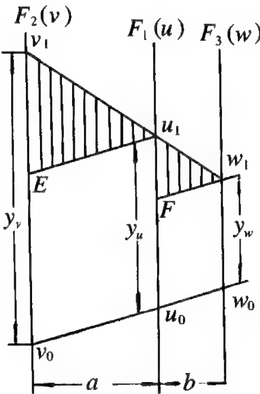


图 1

若选定 $F_2(v)$的图尺系数为 m_v，$F_3(w)$的图尺系数为 m_w；令 $F_2(v)$图尺和 $F_3(w)$图尺与 $F_1(u)$图尺距离之比为 $a:b$，v_1 距 v_0 的实际长度以 y_v 表示，w_1 距 w_0 的实际长度以 y_w 表示，u_1 距 u_0 的实际长度以 y_u 表示：作 $Eu_1 /\!/ Fw_1 /\!/ v_0u_0w_0$；则

$$\triangle v_1Eu_1 \backsim \triangle u_1Fw_1,$$

$$\frac{y_v-y_u}{y_u-y_w}=\frac{a}{b},$$

$$\frac{y_v}{a}+\frac{y_w}{b}=\frac{y_u}{\frac{ab}{a+b}},$$

$$\frac{m_vF_2(v)}{a}+\frac{m_wF_3(w)}{b}=\frac{m_uF_1(u)}{\frac{ab}{a+b}},$$

此式应与$F_1(u)=F_2(v)+F_3(w)$一致，可使

$$m_v=ka,\ m_w=kb,\ m_u=k\frac{ab}{a+b},$$

则
$$\frac{a}{b}=\frac{m_v}{m_w},\ m_u=\frac{m_vm_w}{m_v+m_w}.$$

于是作图时，可根据标值范围、需要的精确度和图幅大小，适当选定 $F_2(v)$的图尺系数 m_v 和 $F_3(w)$的图尺系数 m_w，m_v 与 m_w 选定后，$F_1(u)$的图尺系数就必须是

$$\frac{m_vm_w}{m_v+m_w},$$

各图尺之间的间距应为

$$\frac{a}{b}=\frac{m_v}{m_w}.$$

图 2 是平行图尺计算图的又一示例.

$$5(M+5)^{0.38}=L+0.75P.$$

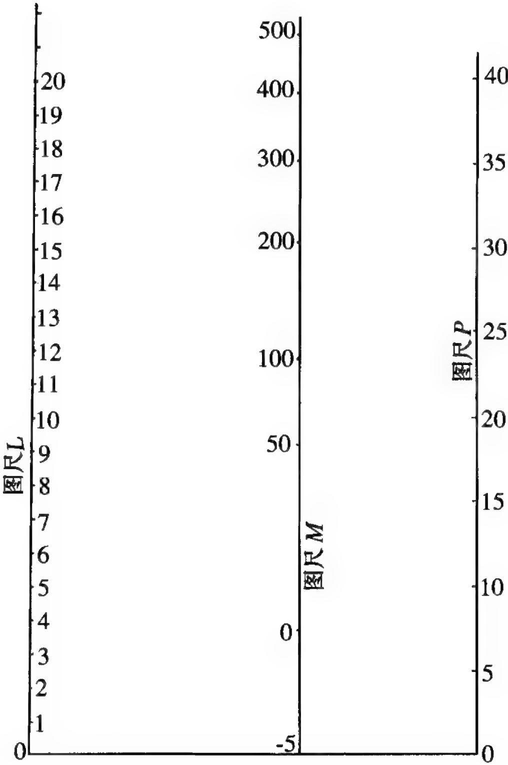


图 2

Z 形图尺计算图(Z chart)　一种常用的计算图. 三个变量的直线图尺成 Z 形，称 Z 形图尺计算图. 适用于方程式 $F_1(u)=F_2(v)\cdot F_3(w)$. 其原理和作法是：如图 1，$A_0$，$B_0$，$C_0$ 为图尺原点，u_1，v_1，w_1

为方程的一组解，以 k 代表 A_0B_0 之长，

$$\triangle u_1A_0w_1 \backsim \triangle v_1B_0w_1,$$

$$\frac{l_u}{l_v}=\frac{l_w}{k-l_w},\ \frac{m_uF_1(u)}{m_vF_2(v)}=\frac{l_w}{k-l_w},$$

但 $F_1(u)/F_2(v)=F_3(w)$，于是

$$F_3(w)=\frac{m_v}{m_u}\cdot\frac{l_w}{k-l_w},$$

由此可得

$$l_w=\frac{k}{\frac{m_v}{m_u}+F_3(w)}\cdot F_3(w),$$

$$m_w=\frac{k}{\frac{m_v}{m_u}+F_3(w)}.$$

作图时可适当选定 m_u 与 m_v 之值，而 m_w 为一变数，作图时可按

$$l_w=\frac{kF_3(w)}{\frac{m_v}{m_u}+F_3(w)}$$

算出各 l_w 的值. 在图中，如以 l_w 代表 w_1B_0 之长，则得

$$l_w=\frac{k}{\frac{m_u}{m_v}F_3(w)+1}.$$

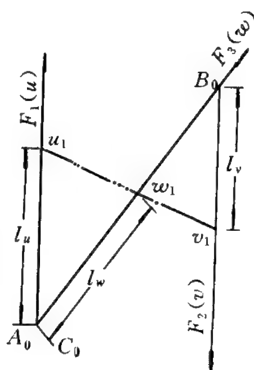


图 1

图 2 为 Z 形计算图的示例. 图示为一两端刚性固定、连续均布载荷的梁，其 A，B 两点反作用力矩大小为 $M_A=M_B=ql^2/12$（M_A 的单位为 kg · cm），式中 q 为均布载荷(kg/cm)，标值范围为 1～8，l 为梁的长度(cm)，标值范围为 100～300.

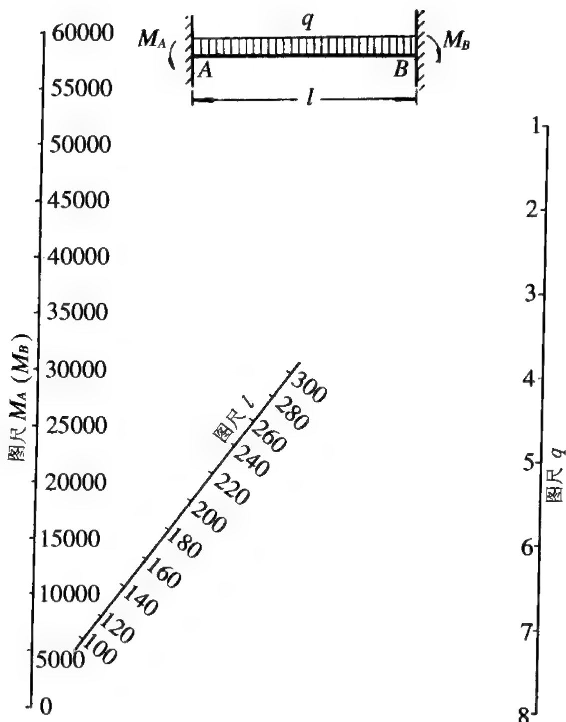


图 2

汇交图尺计算图(concurrent scale chart) 一种常用的计算图. 三个变量的直线图尺交于一点，称汇交图尺计算图，适用于方程

$$\frac{1}{F_1(u)}=\frac{1}{F_2(v)}+\frac{1}{F_3(w)}.$$

如图 1，设在 $F_2(v)$，$F_3(w)$ 图尺上任取 v_1，w_1，连 v_1w_1，在 $F_1(u)$ 图尺上得 u_1，则

$$\frac{1}{F_1(u_1)}=\frac{1}{F_2(v_1)}+\frac{1}{F_3(w_1)},$$

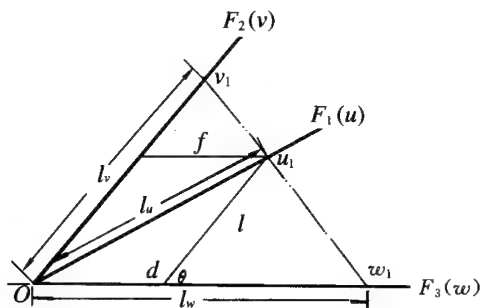


图 1

O 为各图尺的标值起点. 其原理及作法为：由 u_1 作直线 u_1d 平行于图尺 $F_2(v)$，f 平行于图尺 $F_3(w)$，

$$\triangle u_1dw_1 \backsim \triangle v_1Ow_1,\ \frac{l}{l_v}=\frac{l_w-f}{l_w},$$

由此可得
$$\frac{l}{l_v}+\frac{f}{l_w}=1,$$

即
$$\frac{l}{m_vF_2(v)}+\frac{f}{m_wF_3(w)}=1,$$

$$\frac{\frac{l}{m_v}}{F_2(v)}+\frac{\frac{f}{m_w}}{F_3(w)}=1,$$

设
$$\frac{l}{m_v}=F_1(u),\ \frac{f}{m_w}=F_1(u),$$

则
$$\frac{F_1(u)}{F_2(v)}+\frac{F_1(u)}{F_3(w)}=1,$$

即
$$\frac{1}{F_2(v)}+\frac{1}{F_3(w)}=\frac{1}{F_1(u)}.$$

于是 $l=m_vF_1(u)$，$f=m_wF_1(u)$，

$$\frac{l}{f}=\frac{m_v}{m_w}.$$

在 $\triangle Odu_1$ 中，$l_u^2=l^2+f^2+2l\cdot f\cos\theta$，以 $l_u=m_uF_1(u)$，$l=m_vF_1(u)$，$f=m_wF_1(u)$ 代入上式可得 $m_u^2=m_v^2+m_w^2+2m_vm_w\cos\theta$.

这样，作汇交图尺计算图时，m_v，m_w 及 θ 值可适当选定，然后由 $m_v/m_w=l/f$，定出 $F_1(u)$ 图尺的方向，再由 $m_u^2=m_v^2+m_w^2+2m_vm_w\cos\theta$ 求出 m_u 之

值.

图 2 为汇交图尺计算图的示例. 方程式为

$$n=\frac{n_\sigma n_\tau}{\sqrt{n_\sigma{}^2+n_\tau{}^2}},$$

系根据初步确定的结构,用以校核轴的各剖面的疲劳强度极限安全系数,n 为剖面的总安全系数,n_σ 为只考虑弯曲时的安全系数,n_τ 为只考虑扭转时的安全系数.

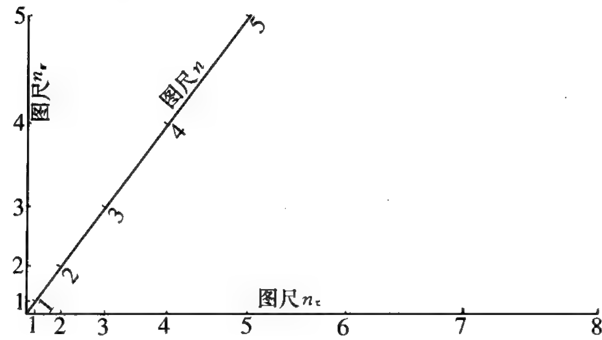


图 2

作计算图时,对有些式子需要反复分析,通过演变,使其符合某一典型方程. 对方程式

$$n=\frac{n_\sigma n_\tau}{\sqrt{n_\sigma^2+n_\tau^2}},$$

进行的演变如下:

$$n^2=\frac{n_\sigma^2 n_\tau^2}{n_\sigma^2+n_\tau^2},$$

$$\frac{1}{n^2}=\frac{n_\sigma^2+n_\tau^2}{n_\sigma^2 n_\tau^2}=\frac{1}{n_\tau^2}+\frac{1}{n_\sigma^2}.$$

比例图尺计算图(proportion chart)　一种常用的计算图. 两两平行的四条直线图尺,称比例图尺计算图,适用于方程

$$\frac{F_1(u)}{F_2(v)}=\frac{F_3(w)}{F_4(s)}.$$

如图 1,图尺夹角可为任意适当角度,一般取为直角,设图尺 $F_1(u)$ 与图尺 $F_2(v)$ 平行,图尺 $F_3(w)$ 与图尺 $F_4(s)$ 平行,若已知 u_1,v_1,w_1,欲求符合方程的 s_1 值,则连 u_1v_1 与 u_0v_0 交于 k,连 w_1k,与图尺 $F_4(s)$ 相交,即得 s_1 值.

其原理及作法如下:

$$\triangle u_1ku_0\backsim\triangle v_1kv_0,\ \frac{l_u}{l_v}=\frac{u_0k}{v_0k},$$

$$\triangle w_1kw_0\backsim\triangle s_1ks_0,\ \frac{l_w}{l_s}=\frac{w_0k}{s_0k},$$

于是
$$\frac{l_u}{l_v}=\frac{l_w}{l_s},\ \frac{m_uF_1(u)}{m_vF_2(v)}=\frac{m_wF_3(w)}{m_sF_4(s)}.$$

因为
$$\frac{F_1(u)}{F_2(v)}=\frac{F_3(w)}{F_4(s)},$$

故
$$\frac{m_u}{m_v}=\frac{m_w}{m_s},$$

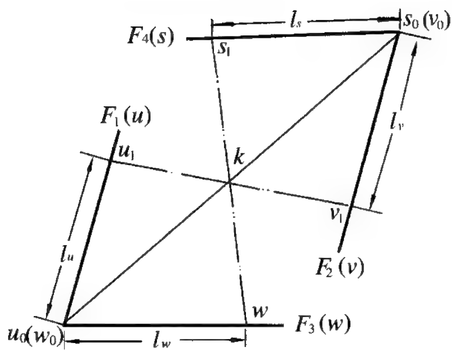


图 1

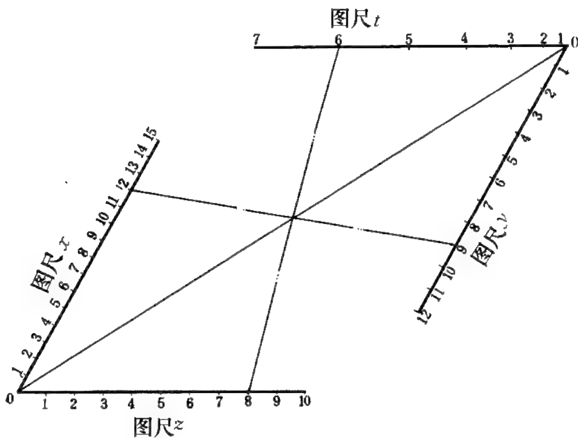


图 2

即四个图尺的图尺系数若符合 $m_u/m_v=m_w/m_s$,即可作成比例图尺计算图. 图 2 为比例图尺计算图的示例,图中设选定 $m_x=6,m_y=4,m_z=3$,由 $m_x/m_y=m_z/m_t$,可算出 $m_t=2$.

$$\frac{x}{2y}=\frac{3z}{t^2}.$$

有一条曲线图尺的计算图(chart with one curved scale)　一种常用的计算图. 若方程为

$$F_1(u)+F_2(v)\cdot F_3(w)=F_4(w),$$

即式中两次出现 w 的某函数,一是单独的,一是与另一其他函数相乘,则可画成由 u,v 两条平行图尺和 w 曲线图尺组成的计算图. 如图 1,若在 v 图尺上取 v_1,w 图尺上取 w_1,连 v_1w_1 与 u 图尺交于 u_1,则 u_1,v_1,w_1 为方程的一组解. 其原理如下:作 w_1M,v_1N 平行于标值基线 v_0u_0,于是

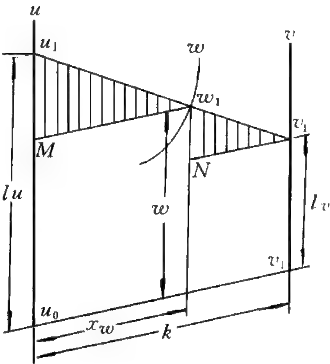


图 1

$$\triangle u_1Mw_1\backsim\triangle w_1Nv_1,$$

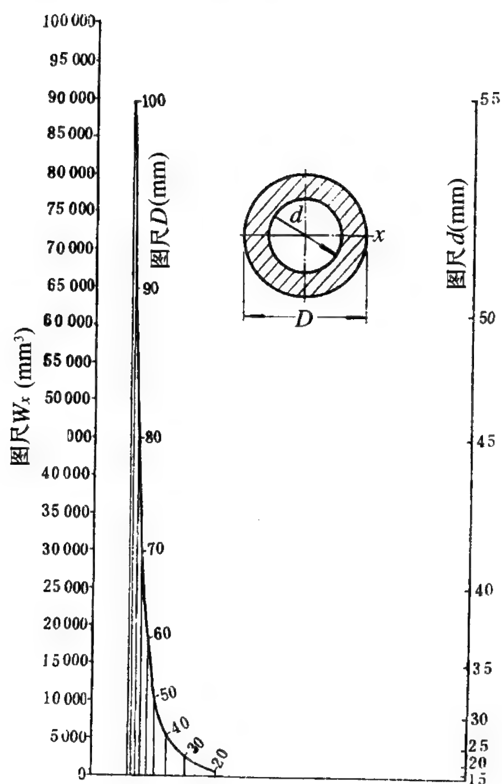


图 2

$$\frac{l_u-y_w}{y_w-l_v}=\frac{x_w}{k-x_w},$$

$$l_u+l_v\left(\frac{x_w}{k-x_w}\right)=\frac{ky_w}{k-x_w},$$

$$m_uF_1(u)+m_vF_2(v)\frac{x_w}{k-x_w}=\frac{ky_w}{k-x_w},$$

$$F_1(u)+F_2(v)\frac{m_v}{m_u}\cdot\frac{x_w}{k-x_w}=\frac{ky_w}{(k-x_w)m_u},$$

此应与 $F_1(u)+F_2(v)\cdot F_3(w)=F_4(w)$一致. 则需

$$\frac{m_v}{m_u}\cdot\frac{x_w}{k-x_w}=F_3(w),$$

$$\frac{ky_w}{(k-x_w)m_u}=F_4(w),$$

由此可得：

$$x_w=\frac{km_uF_3(w)}{m_uF_3(w)+m_v},\ y_w=\frac{m_um_vF_4(w)}{m_uF_3(w)+m_v}.$$

图 2 是有一条曲线图尺的计算图的示例. 方程为

$$W_x=\frac{\pi D^3}{32}\left(1-\frac{d^4}{D^4}\right),$$

式中 W_x 为轴 x 的抗弯剖面系数. D 与 d 为剖面的外径与内径.

有两条曲线图尺的计算图(chart with two curved scales) 一种常用的计算图. 方程式中有一个 u 的函数、两个 v 的函数和两个 w 的函数，其关系式为

$$F_2(u)=\frac{F_3(v)F_6(w)-F_5(w)F_4(v)}{F_3(v)-F_5(w)},$$

则可画成由一条直线图尺和两条曲线图尺组成的计算图. 其原理如下：如图 1，u 可取为直线图尺，v，w 为曲线图尺，v 图尺与 w 图尺上的标值点需以坐标表示. 设已知 u_1，v_1，连 u_1v_1，在 w 图尺上得 w_1，则 u_1，v_1，w_1 为方程的一组解. 设 $x_u=m_uF_1(u)=0$，

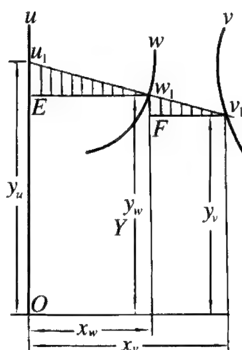


图 1

$y_u=m'_uF_2(u)$，$x_v=m_vF_3(v)$，

$y_v=m'_vF_4(v)$，$x_w=m_wF_5(w)$，

$y_w=m_{w'}F_6(w)$，

$$\triangle u_1Ew_1\backsim\triangle w_1Fv_1,\ \frac{y_u-y_w}{y_w-y_v}=\frac{x_w}{x_v-x_w},$$

化简得

$$y_u=\frac{x_vy_w-x_wy_v}{x_v-x_w},$$

$$m'_uF_2(u)=\frac{m_vF_3(v)m'_wF_6(w)-m_wF_5(w)m'_vF_4(v)}{m_vF_3(v)-m_wF_5(w)}.$$

此式应与

$$F_2(u)=\frac{F_3(v)F_6(w)-F_5(w)F_4(v)}{F_3(v)-F_5(w)}$$

一致，则需 $m_u=m_v=m_w$ 及 $m'_u=m'_v=m'_w$.

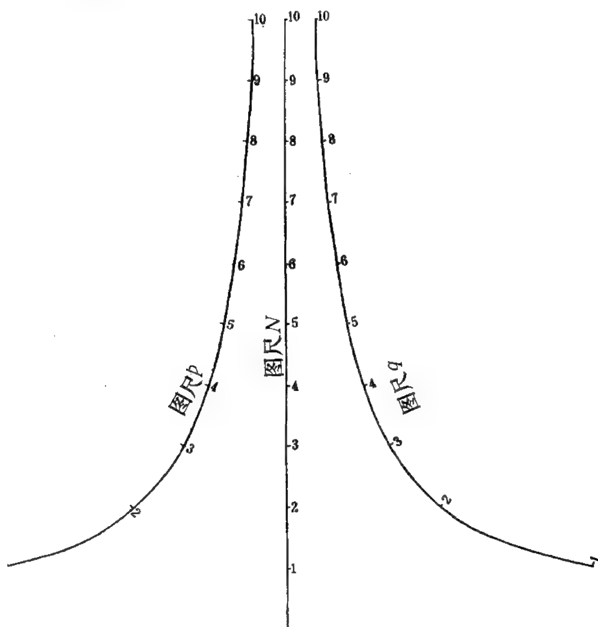


图 2

图 2 是有两条曲线图尺的计算图的示例. 方程为

$$N = \frac{p^2 + q^2}{p + q}.$$

首先进行观察探索，方程可演变为

$$N = \frac{\left(-\frac{1}{p}\right)q - p\left(\frac{1}{q}\right)}{\left(-\frac{1}{p}\right) - \left(\frac{1}{q}\right)}.$$

有圆形图尺的计算图（chart with circular scale）　一种常用的计算图. 当方程为 $F_1(u) =$

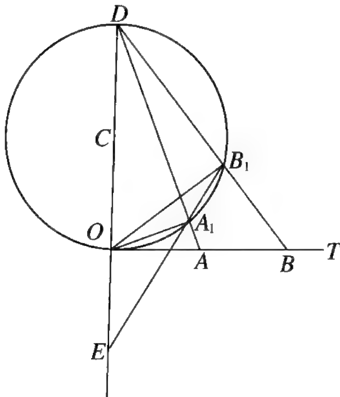


图 1

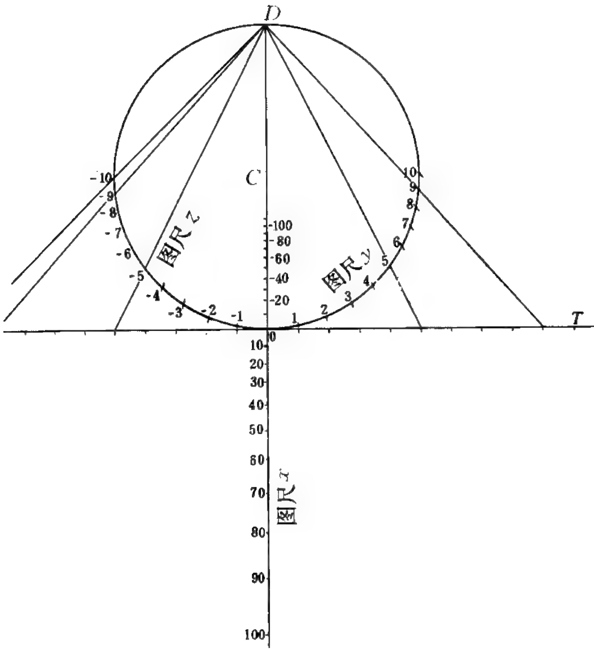


图 2

$F_2(v) \cdot F_3(w)$时，可以画成Z形图尺计算图、平行图尺计算图，也可以画成有圆形图尺的计算图. 其作法及原理如下所述：如图1，设圆的直径 $OD=d$，OT 为圆的切线，在 OT 上任取两点 A，B，设 $OA=a$，$OB=b$，连 AD，与圆交于 A_1，连 BD，与圆交于 B_1，连 B_1A_1 与 DO 交于 E，则 E 点的位置由乘积 ab 决定，当乘积 ab 不变时，E 点位置不变. 这是由于 $\triangle OAA_1 \backsim \triangle ODA_1$，$\triangle OBB_1 \backsim ODB_1$，$\triangle EOA_1 \backsim$

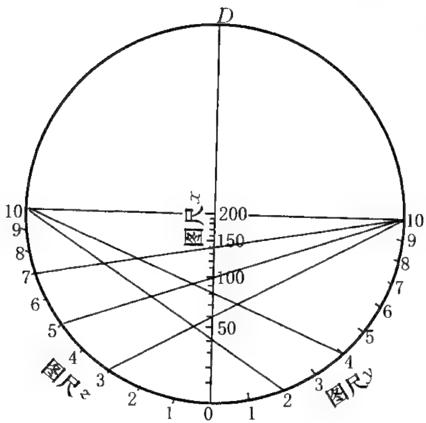


图 3

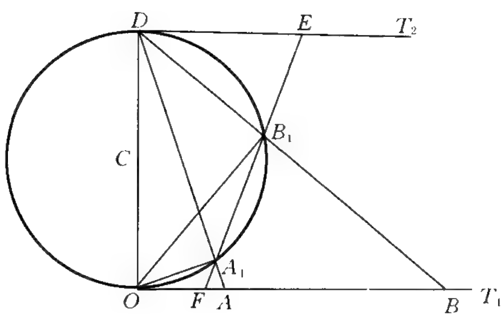


图 4

$\triangle EDB_1$，$\triangle EOB_1 \backsim \triangle EDA_1$，$\triangle EOA_1 \backsim \triangle EDB_1$，可得

$$EO = \frac{abd}{d^2 - ab},$$

可知若 d 之值和乘积 ab 之值不变，则 E 点位置不动.

图2是 $x=2yz$ 的有圆形图尺计算图. 若变更图尺的正负方向，则伸出圆周以外的图尺 x 可转换到圆周以内，如图3.

当方程为$F_1(u)=F_2(v)+F_3(w)$时，也可以画成有圆形图尺的计算图，其原理为：如图4，设圆的直径 $OD=d$，过 O，D 作圆的切线 OT_1，DT_2. 在 OT_1 上任取两点 A，B，设 $OA=a$，$OB=b$，连 DA，DB，在圆上得 A_1，B_1 两点，连 A_1B_1，与两切线分别交于 F，E，则 E 点的位置由 a，b 之和决定，即如果 $a+b$ 之值不变，则 E 点的位置不变. 通过 $\triangle OAA_1 \backsim \triangle OAD$，$\triangle OA_1D \backsim \triangle OAD$，$\triangle FA_1A \backsim EA_1D$，即可证明

$$DE = \frac{d^2}{a+b},$$

即 DE 之值由 $a+b$ 决定. 图5是 $x^2+mx+n=0$ 的有圆形图尺的计算图. 设有方程 $x^2+14x+24=0$，欲求其两根 x_1，x_2，在 m 图尺上找到14，在 n 图尺上找到24，以直线连此两标值点，与圆周交于-12，$-$

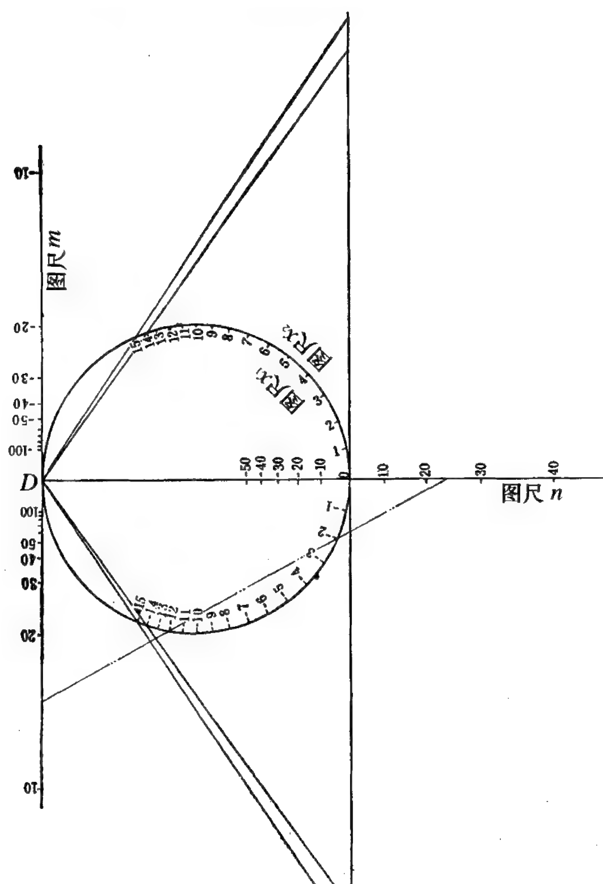


图 5

2,即为方程的两根.

有网线图尺的计算图(chart with net scale) 一种常用的计算图.若方程为

$$F_1(u)F_3(w)+F_2(v)F_4(w)+F_5(t)=0,$$

可以画成两条平行图尺和一个网线图尺的计算图.其作法和原理如下所述:如图 1,$F_1(u)$和$F_2(v)$为两条平行图尺,设两图尺间距为k,w与t组成一组网线图尺.以标值基线AB为x轴,过AB中点平行于$F_1(u)$,$F_2(v)$两图尺的直线为y轴.在方程中,若$F_1(u)=0$,则

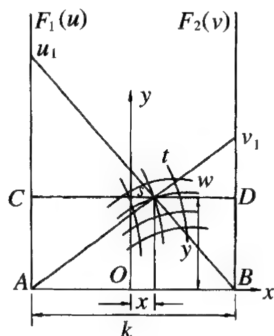


图 1

$$F_2(v)=\frac{-F_5(t)}{F_4(w)}=Bv_1.$$

若$F_2(v)=0$,则

$$F_1(u)=\frac{-F_5(t)}{F_3(w)}=Au_1.$$

Av_1与Bu_1相交于$s(x,y)$点,s为网线上一点,是由w和t两个变量的数值所决定,当w和t给出一系列数值时,即可得网线图尺.

过s作$CD/\!/AB$,通过$\triangle u_1Cs\backsim\triangle u_1AB$,$\triangle v_1Ds\backsim\triangle v_1BA$,可导出

$$x=\frac{k}{2}\cdot\frac{F_4(w)-F_3(w)}{F_4(w)+F_3(w)},$$

$$y=\frac{-F_5(t)}{F_4(w)+F_3(w)}.$$

设$F_1(u)$与$F_2(v)$的图尺系数为m_u,m_v,可继续演变为

$$\begin{cases}x=\dfrac{k}{2}\cdot\dfrac{m_uF_4(w)-m_vF_3(w)}{m_uF_4(w)+m_vF_3(w)},\\ y=\dfrac{-m_um_vF_5(t)}{m_uF_4(w)+m_vF_3(w)}.\end{cases}$$

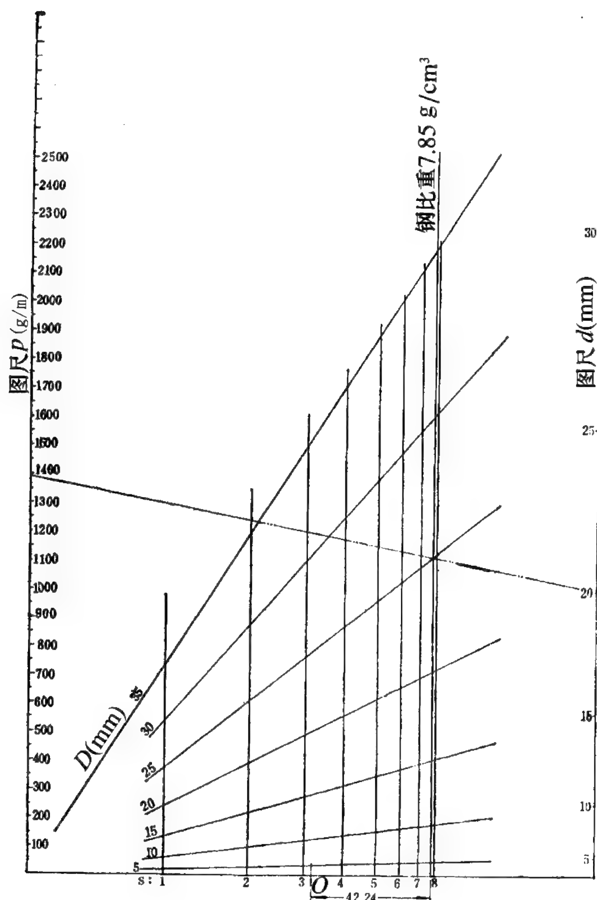


图 2

图 2 为有网线图尺计算图的示例.方程为

$$p=S\left(\frac{\pi D^2-\pi d^2}{4}\right),$$

式中p为每米长的圆管重量(g/m),D为圆管外径(mm),d为内径(mm),S为材料比重(g/cm³).为使网线在图中布置匀称,将原方程两端除以 10 变为

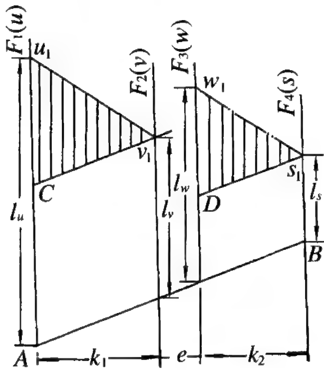
$$\frac{p}{10}=\frac{S}{10}\left(\frac{\pi D^2-\pi d^2}{4}\right).$$

图 2 中双点划线表示：钢材比重为 7.85 克/c.c.，钢管外径 $D=25\text{mm}$，内径 $d=20\text{mm}$，每米长的钢管重量为 1390g，即 1.39kg.

平行指示线计算图（parallel index lines chart） 一种常用的计算图. 对于方程

$$F_1(u)+F_2(v)=F_3(w)+F_4(s),$$

可作四条平行的直线图尺，而求解的指示线互相平行，称平行指示线计算图.



（e 的大小可任意给定）

图 1

如图 1，设 u_1,v_1,w_1,s_1 为方程的一组解，$u_1v_1 /\!/ w_1s_1$，作 Cv_1 和 Ds_1 分别平行于标值基线 AB，则

$$\triangle u_1Cv_1 \backsim \triangle w_1Ds_1,\ \frac{l_u-l_v}{k_1}=\frac{l_w-l_s}{k_2},$$

$$\frac{m_uF_1(u)-m_vF_2(v)}{k_1}=\frac{m_wF_3(w)-m_sF_4(s)}{k_2}.$$

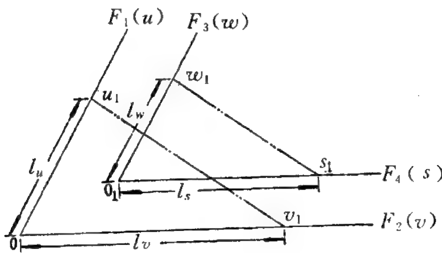


图 2

欲使符合 $F_1(u)+F_2(v)=F_3(w)+F_4(s)$，则需

$$m_u=-m_v,\ m_w=-m_s,\ \frac{m_u}{m_w}=\frac{k_1}{k_2}.$$

对于方程

$$\frac{F_1(u)}{F_2(v)}=\frac{F_3(w)}{F_4(s)}$$

的比例图尺计算图也可作成两两平行的四条直线图尺，而求解的指示线互相平行（如图 2）.

图 3 是平行指示线计算图的示例，方程为

$$d_{分}=\frac{m_nz}{\cos\beta},$$

式中 $d_{分}$ 为斜齿圆柱齿轮分度圆直径，m_n 为法向模数，z 为齿数，β 为螺旋角.

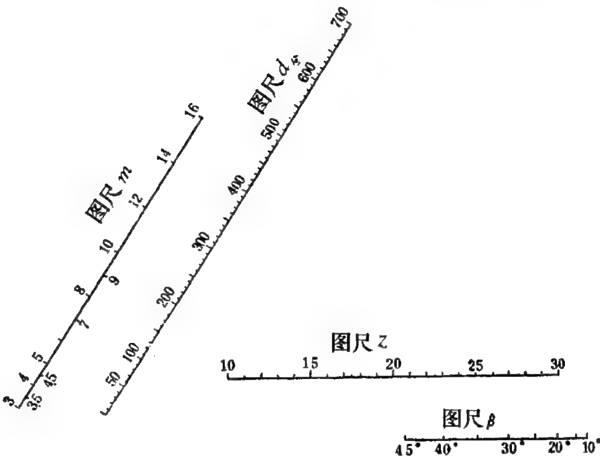


图 3

垂直指示线计算图（orthogonal index lines chart） 一种常用的计算图. 对方程

$$F_1{}^2(u)+F_2(v)F_1(u)+F_3(w)=0,$$

可作成两条平行直线图尺和一条与之垂直的直线图尺，而求解的指示线是互相垂直的两直线，称垂直指示线计算图. 如图 1，设指示线在计算图上截出 u_1，v_1，w_1 一组解，则

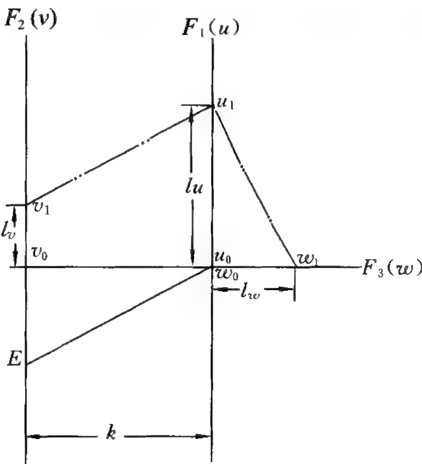


图 1

$$\triangle u_1u_0w_1 \backsim \triangle u_0Ev_0 \quad (v_1u_1 \perp u_1w_1),$$

$$\frac{l_u}{l_w}=\frac{k}{l_u-l_v},$$

可得

$$l_u{}^2-l_ul_v-kl_w=0,$$

$$m_u{}^2F_1^2(u)-m_uF_1(u)m_vF_2(v)-km_wF_3(w)=0,$$

此式应与

$$F_1{}^2(u)+F_1(u)F_2(v)+F_3(w)=0$$

一致，则需使 $-m_v/m_u=1$，即 $m_u=-m_v$，$-km_w/m_u{}^2=1$，即 $m_w=-m_u^2/k$. 使用垂直指示线计算图时，可

在透明板或透明纸上画两条互相垂直的直线作为工具.

图 2 为垂直指示线计算图的示例. 方程为

$$x^2+mx+n=0.$$

图中双点划线表示:

1. 当 $x=2, n=-10$ 时, 求得 $m=3$;

2. 若已知 $m=3, n=-10$ 时, 则 x 有两解, 即 $x_1=2, x_2=-5$.

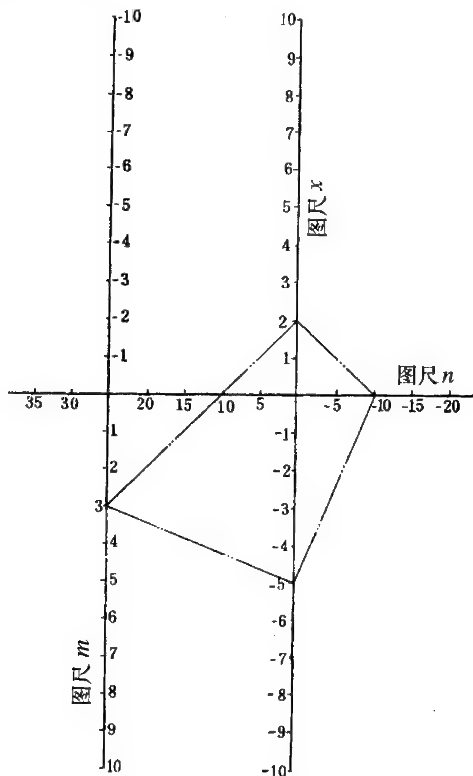
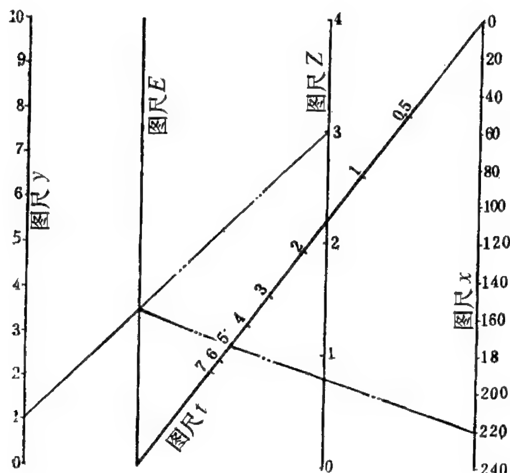


图 2

复合计算图(combination chart) 一种常用的计算图. 由对已知方程作成两种或多种形式计算图组合而成的计算图, $x=(2y+3z)4t$ 称为复合计算图. 图示是由一个平行图尺计算图和一个 Z 形图尺计算图组合而成的复合计算图, 图尺 E 称辅助图尺, 其上不需标值.



圆弧指示线计算图(arc indexline chart) 一种常用的计算图. 对方程

$$F_1(u)=F_2(v)+F_3(w),$$

可作成相交于一点的三条直线图尺, 而求解的指示线是圆弧. 圆弧指示线计算图的图尺布置如图 1 所示, O 为各图尺标值起点, $\angle v_1Ou_1$ 与 $\angle u_1Ow_1$ 各为 $60°$, $u_1v_1w_1$ 为等边三角形(u_1, v_1, w_1 为方程的一组解). 若已知 v_1, w_1, 则以 v_1 或 w_1 为圆心、v_1w_1 为半径画圆弧, 即可在 u 图尺上得到 u_1.

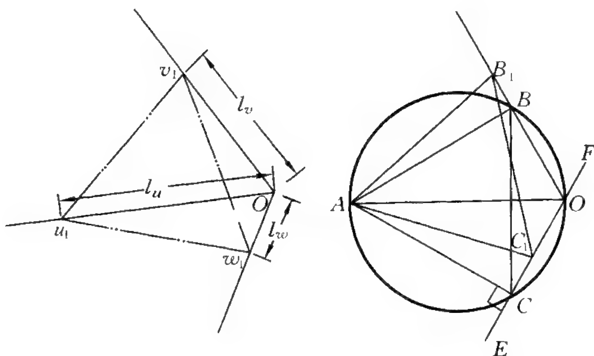


图 1　　图 2

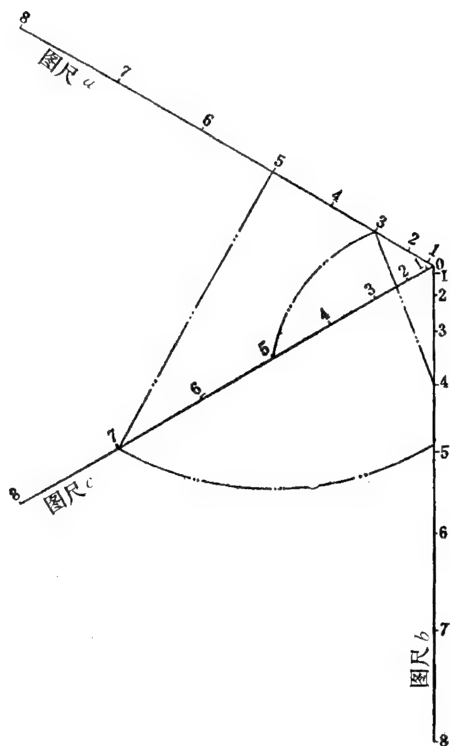


图 3

其原理如下: 如图 2, 设 EF 为一定直线, 自任

意一点 A 作 AC 垂直于 EF，以 AC 为边，作正三角形 ABC，在 EF 上任取一点 C_1，以 AC_1 为边，作另一正三角形 AB_1C_1，连 B_1B 与 EF 交于 O，则 $\triangle ABB_1 \backsim \triangle ACC_1$，$\angle ABB_1 = \angle ACC_1 = 90°$，$A,B,O,C$ 四点共圆，$\angle BOC = 120°$，而 OA 为 $\angle BOC$ 的角平分线. 这样的情况说明 C_1 沿定直线移动时，B_1 沿过 B 点垂直于 AB 的定直线 BO 移动.

如图 1，在 $\triangle u_1Ow_1$ 中，$u_1w_1^2 = l_u^2 + l_w^2 - 2l_ul_w \cdot \cos 60°$；在 $\triangle v_1Ow_1$ 中，$v_1w_1^2 = l_v^2 + l_w^2 - 2l_vl_w \cos 120°$；$u_1w_1 = v_1w_1$，可得 $l_u = l_v + l_w$，

$$m_uF_1(u) = m_vF_2(v) + m_wF_3(w),$$

此式应与 $F_1(u) = F_2(v) + F_3(w)$ 一致，则需 $m_u = m_v = m_w$.

图 3 是圆弧指示线计算图的示例. 方程是 $a^2 + b^2 = c^2$. 图中双点划线表示：$a=3, b=4$ 时求得 $c=5$；$c=7, a=5$ 时，求得 $b=4.9$.

共点计算图与共线计算图的关系(relationship between concurrency and alignment nomograms) 叙述共点计算图与共线计算图的基本关系，其间的关系用对偶定理表明. 该定理指出："共点计算图的

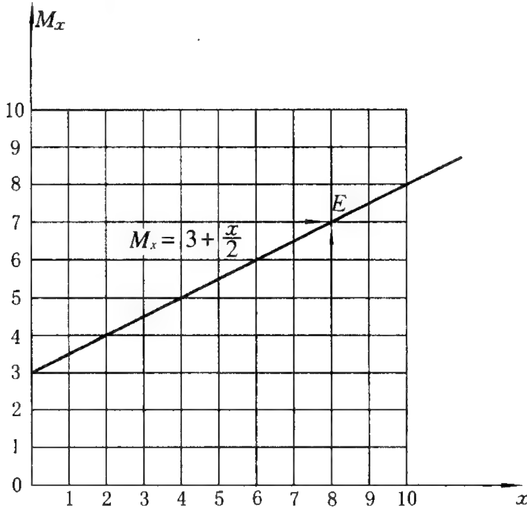


图 1

每一条线相当于共线计算图的一个对应点."如图 1，设方程为

$$M_x = 3 + \frac{x}{2},$$

在共点计算图上，线上任一点 E 为三条线的交点，即 $x=8, M_x=7$ 和过点 $(0,3)$，$(8,7)$ 的直线.

$$M_x = 3 + \frac{x}{2}$$

作成共线计算图，如图 2，通过 $x=0, M_x=3$ 的直线 AC 和通过 $x=8, M_x=7$ 的直线 BD 相交于 P. 则所有满足方程

$$M_x = 3 + \frac{x}{2}$$

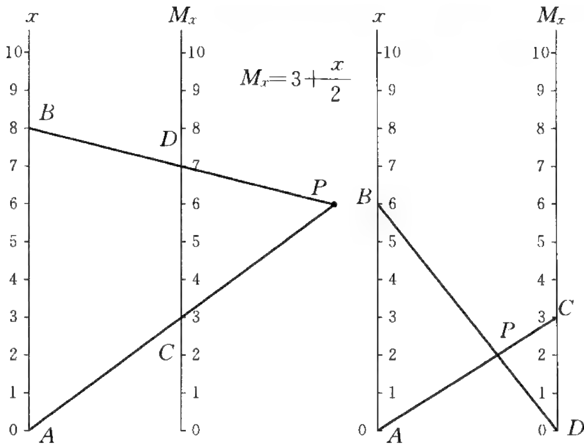


图 2　　图 3

的直线都通过 P. 于是共点计算图上直线 $x=8$ 对应于共线计算图上图尺 x 上的 B 点，共点计算图上直线 $M_x=7$ 对应于共线计算图上的 D 点，共点计算图上直线

$$M_x = 3 + \frac{x}{2}$$

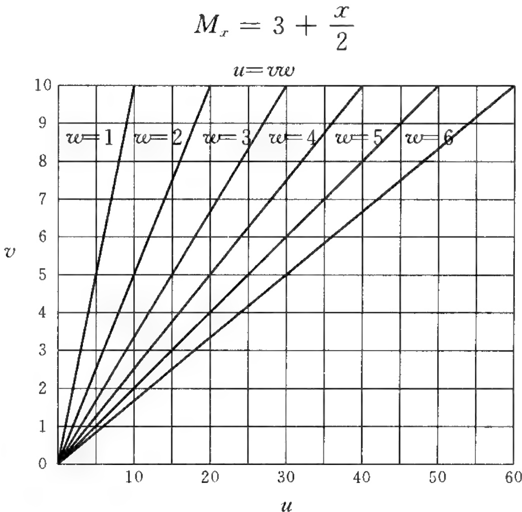


图 4

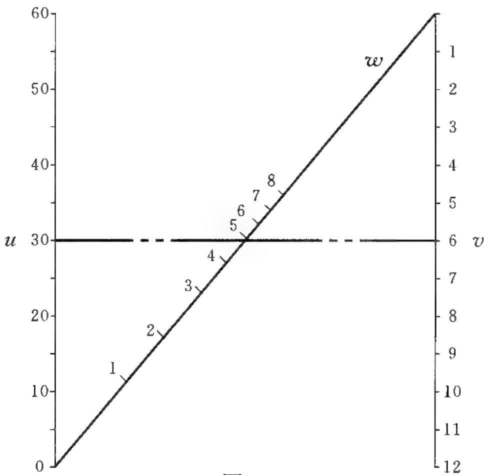


图 5

对应于共线计算图上的 P 点. 令

$$\frac{CP}{AP} = \frac{CD}{AB} = m,$$

若 m 为负值(例如方程

$$M_x=-\frac{x}{2}+3,$$

m 为 $-1/2$),则 P 的位置在图尺 M_x 与图尺 x 中间,如图 3. 若 x 与 M_x 标值方向相反,而 m 为正,P 也在图尺 M_x 与图尺 x 的中间.

由以上可知,共点计算图与共线计算图可以互换. 图 4、图 5 是共点计算图变换为共线计算图的又一示例.

对偶定理(law of duality) 见“共点计算图与共线计算图的关系”.

曲线图尺的校直(rectification of curve scale) 绘制共线计算图的一种方法. 为一个方程式作共点计算图,往往是画出曲线,应用对偶定理将共点计算

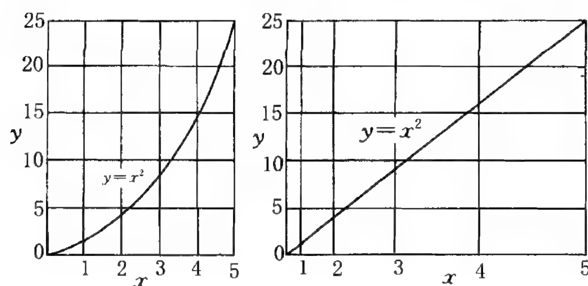


图 1　　图 2

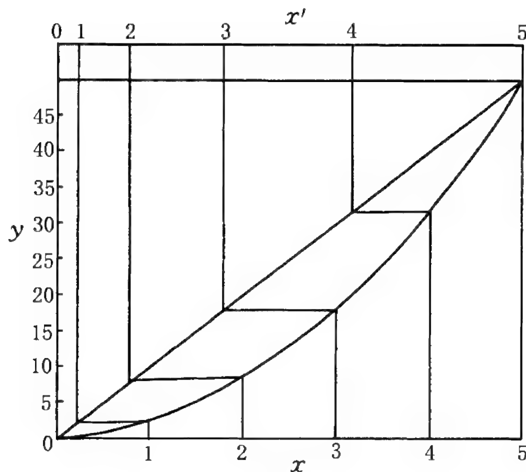


图 3

图变成共线计算图时,最好是先将曲线校直. 在作科学实验或进行调查统计时,将所得众多数据在坐标系中标出,而连成的大多数是曲线,将这样的实验与统计曲线校成直线,可以更好地绘出共线计算图. 图 1 是为 $y=x^2$ 做出的共点计算图. 如果 x 坐标轴按 x^2 标值,则曲线就成了直线,如图 2. 图 3 是未知方程的共点计算图,经校直后(x 坐标轴变成 x' 坐标轴),曲线就变成了直线.

多曲线图尺的校直(rectification of multi-curve scale) 制作共线计算图的一种方法. 在共点计算图中,如果曲线图尺是由两条或多条曲线组成,则要校直为两条或多条直线,然后再作共线计算图. 如图 1,左上为已知共点计算图,坐标轴为 x,y,在两曲线

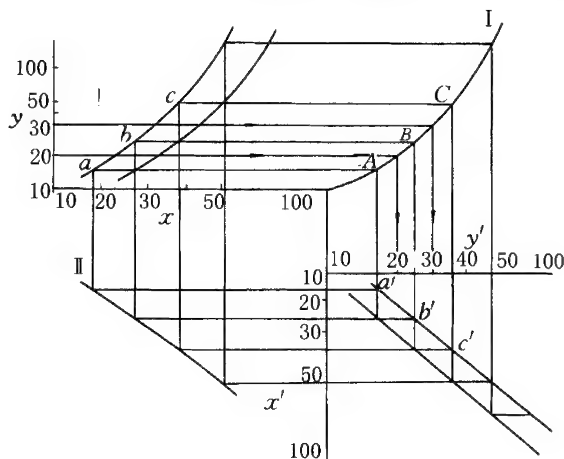


图 1

间作梯级状线 $abc\cdots$,在图的右下,另设两坐标轴 x',y',于适当位置设置两条平行直线,两平行线间作相应的梯级状线 $a'b'c'\cdots$,自 a 作水平线,自 a' 作铅直线,两线相交于一点 A,自 b 作水平线,自 b' 作铅直线,两线相交于一点 B,$\cdots$,A,B,$\cdots$各点连成的曲线 I 称转换曲线. 同样作法,Ⅱ 也是转换曲线. 通过转换曲线 Ⅰ,Ⅱ,将 y 轴上的标值刻度移到新轴 y' 上,x 轴上的标值刻度移到新轴 x' 上,在 x',y' 两轴组成的坐标系中,两曲线图尺即被校成直线.

对于三条以上的一族曲线,能否在同一新坐标轴体系 x',y' 中校直,需要作“双线性试验”. 如能通过双线性试验,则可确定这些曲线可在同一 x',y' 坐标轴体系中校直. 所谓通过双线性试验,即如图 2 中的矩形链能形成一个闭合环. 要在曲线族之间适当任定几处或多处作这样的试验.

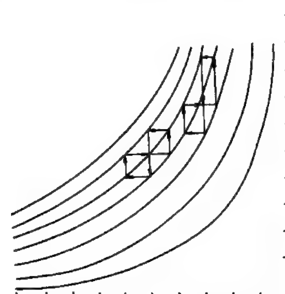


图 2

转换曲线(transformation curve) 见“多曲线图尺的校直”.

双线性试验(bilinearity test) 见“多曲线图尺的校直”.

撰　稿　杜喜清　李仙梅　沈肇敏　张尔其　高　健　梁世熙

审　阅　石光源　洪钟德　梁世熙

计　算　机　科　学

计算机科学(computer science)　研究计算机本身及其实现方法的科学.计算机是一种现代化的信息处理工具.计算机科学是研究计算机本身和利用计算机进行信息获取、信息表示、信息储存、信息处理、信息控制等的理论、原则和方法的学科.早在现代计算机问世前一百多年,人们就在不断探索理想计算装置的原理、结构和实现方法.20 世纪 40 年代,由于电子技术和计算理论的重大进展,电子数字计算机应运而生,计算机科学也随之发展起来.计算机科学主要包括理论计算机科学、计算机体系结构、计算机软件和人工智能等内容.

理论计算机科学也称为计算机科学理论.主要包括计算机科学的数学基础、计算理论和程序理论三部分.其中,计算理论和程序理论是计算机科学理论的两大支柱.计算机科学的数学基础以离散数学为核心.离散数学是以离散结构为主要研究对象的现代数学分支.通常应包括集合论、图论、组合学、数理逻辑、证明技术、离散概率论等.计算理论是研究计算和计算机械的数学理论.主要包括算法、算法学、计算复杂性理论、可计算性理论、自动机理论、形式语言理论、分布式算法、高级算法分析、加密算法、几何算法等.程序理论是研究程序的语义性质和程序的设计及开发方法的理论.主要包括程序语义理论、数据类型理论、程序逻辑理论、程序验证理论、开发程序设计理论、事件驱动与并发程序设计、以及高级 API 使用等.

计算机体系结构着重研究计算机系统的物理或硬件结构、各组成部分的属性以及这些部分的相互联系.它包括计算机类型、计算机组织、计算机系统结构、计算机网络、计算机 RAS 技术和计算机性能评价.计算机类型,按计算机内数据表示的方式可分为数字计算机、模拟计算机、混合计算机;按系统规模和性能可分为微型计算机、小型计算机、大型计算机、巨型计算机;按用途可分为通用计算机、专用计算机等.计算机组织主要研究计算机运算基础、算术逻辑运算、指令系统、中央处理器、存储器组织、输入输出技术等.计算机系统结构主要研究处理器体系结构、存储系统、并行处理系统、分布式处理系统、开放系统等.计算机网络主要研究网络体系结构、网络协议、网络互联技术、网络分类、以网络为中心的计算、客户机服务器模式、网络管理、因特网等.计算机 RAS 技术主要研究计算机系统的可靠性、可维护性、可用性等.计算机性能评价主要研究计算机安全、系统维护、系统性能指标、运算速度评价等.

计算机软件一般指计算机系统中的程序及其文档.作为学科,计算机软件是研究软件在研究、开发、维护和使用过程中所涉及的理论、方法和技术.其内容包括程序、软件语言、软件方法、软件系统、软件工程等.程序是计算任务的处理对象和处理规则的描述,研究内容包括程序的构成、数据类型、数据结构等.软件语言包括程序设计语言、软件需求定义语言、功能性语言、设计性语言、文档语言等.软件方法学主要研究程序设计方法、软件自动化方法、自顶向下方法、自底向上方法、面向对象方法、模块化方法等.软件系统主要研究操作系统、编译系统、数据库系统、分布式软件系统、人机交互系统等.软件工程主要研究软件结构、软件开发方法、软件生存周期、软件开发模型、软件过程、软件工具、软件工程环境、软件开发环境、计算机辅助软件工程、软件质量、软件安全性等.

人工智能是研究解释和模拟人类智能、智能行为及其规律的一门学科.其主要任务是建立智能信息处理理论,进而设计并实现可以展现某些近似于人类智能行为的计算系统.研究内容包括知识工程、模式识别、机器学习、自然语言处理、神经计算、智能机器人等.知识工程主要研究知识表示、自动推理、专家系统、启发式搜索等.模式识别主要研究模式识别方法、计算机视觉等.机器学习主要研究归纳学习、类比学习、分析学习、遗传学习、知识发现等.自然语言处理主要研究自然语言理解、机器翻译等.神经计算主要研究人工神经网络、感知器、联想记忆、反传学习、神经网络模型等.

计算机科学是一门年轻的学科,但它发展迅速、影响巨大.计算机系统已渗透到国民经济的各个领域和人民生活的各个方面,成为科技进步的重要标志及社会信息化的重要组成部分.它在各行各业的广泛应用,产生了显著的经济效益和社会效益,引起了产业结构、经营管理和服务方式等的巨大变革.计算机科学与其他学科的结合,改进并发展了其他学科的研究工具和研究方法,促进了其他学科的发展和新型学科的诞生.与此同时,这种结合还诱发出大量的不断增长的对其自身的社会需要,成为推动自身前进的巨大动力.当今,对计算机科学与技术的研究水平,对计算机系统的开发能力、生产规模和应用程度已成为衡量一个国家现代化水准和国力的重要标志.此外,人们还应关注因特网对社会的冲击,关

注计算机犯罪、网络经济及其引发的社会道德等问题.

计算机科学的数学基础

计算机科学的数学基础(mathematical foundation for computer science) 在计算机科学的研究与应用中所需要的数学基础知识的总称.由于计算机科学技术的飞速发展,它的应用已深入到社会的各个领域,从这个意义上来说,计算机科学的数学基础是非常广泛的,应包括数学的一切分支.从传统观念来说,它主要包括集合论、图论、近世代数、数理逻辑、组合学等数学分支.从实际应用的角度来说,数论、概率论、统计学、信息论、几何学、拓扑学、数学物理方法、计算数学等分支都是重要的数学基础.

离散数学(discrete mathematics) 见《数学辞海》第四卷同名条.

组合学(combinatorics) 见《数学辞海》第四卷同名条.

形式逻辑(formal logic) 见《数学辞海》第一卷同名条.

算法(algorithm) 解题过程的精确描述.它是一个有限规则的集合,这些规则确定了求解某一类问题的一个运算序列,对于某一类问题的任何初始输入,它能机械地一步一步地计算,并且通过有限步骤之后,计算终止并产生输出.一个算法有以下五个特性:

1. 有限性.一个算法必须总是在执行有限步之后结束.

2. 确定性.算法的每一个步骤必须是确切地定义的.

3. 输入.一个算法有 0 个或多个输入.

4. 输出.一个算法有 1 个或多个输出.

5. 能行性.算法中要执行的每一个计算步骤都是可以在有限时间内完成的.

算法的严格定义,可以用图灵机来形式地描述.按照丘奇-图灵论题,任意一个算法都可以用一个图灵机来实现,反之,任意一个图灵机都表示一个算法.算法按其所处理的对象的不同,可分为数值算法和非数值算法两大领域.数值算法主要包括多项式计算、矩阵计算、有限元计算、数论计算等有关数值计算的算法问题;非数值算法主要包括分类、搜索、几何问题的计算、离散结构的计算、模式匹配等有关非数值计算的算法问题.按照处理问题的方式进行分类,大体上又可分为串行算法和并行算法两大类.

20 世纪 60 年代以来,随着各个领域算法研究工作的发展,产生了一个崭新的学科,这就是算法学.算法学系统地研究算法的设计、分析与验证.算法设计的任务是对各类具体的问题设计高质量的算法,并研究设计算法的一般规律和方法.算法分析研究算法的复杂性、探讨某种具体算法适合于哪类问题以及研究某种具体问题宜于采用某类算法等.验证的目的是证明算法的正确性.

算法设计(design of algorithm) 研究设计算法的一般规律和方法.算法不仅可以由人,而且可以由计算机来完成.用计算机解决问题的过程大致可以分成三个阶段:分析问题、设计算法和实现算法.

1. 分析问题.包括问题的陈述和模型的确定.为了设计求解某一问题的算法,设计者首先必须分析问题,并用科学的语言把问题的所有已知条件和需要的答案陈述清楚,然后合理地选择或拟定描述问题的数学模型.

2. 设计算法.包括算法设计和正确性证明.算法设计是一种复杂的创造性劳动,它要求设计者充分发挥主观能动性,充分运用各种知识和抽象思维,逐渐形成算法的基本思想,勾画出一个算法的各具体步骤.算法的正确性证明与程序的正确性证明一样,是算法理论的重要内容之一(参见“程序验证”).

3. 实现算法.把一个算法正确地编制成一个机器程序,并在计算机上测试和运行.

常用的算法设计方法主要有分治法、贪婪法、回溯法等.

分治法(divide-and-conquer algorithm) 常用的一种算法设计方法.其基本思想是把一个问题分解成若干个子问题,然后由子问题的解得出原来问题的解.在解决一个输入量为 n 的大规模问题时,如果能把问题分割成 k 个子问题,$1<k\leqslant n$,这些子问题都是可解的,并且能由这些子问题的解组合成原来问题的解,那么这种分治法便是可行的.由分治法产生的子问题往往是原问题的较小模式.在这种情况下,反复应用分治策略,可以使子问题与原问题类型一致而规模却越来越小,最终使子问题缩小到无需再分就足以解决.这样,一般会导致递归过程的产生.因此,分治和递归像一对孪生兄弟一样,经常同时出现在那些有效的算法之中.分治和递归技术的使用,还可以降低一些算法的时间复杂度.例如,采用分治和递归技术的两个具有 n 项系数多项式乘法,时间复杂度由通常的$O(n^2)$降为$O(n^{1.59})$;n 阶矩阵的 Strassen 乘法,由通常的$O(n^3)$降为 $O(n^{2.81})$;n 个元素的归并分类法和快速分类法只需要 $O(n\log_2 n)$次比较;n 维的快速傅里叶变换只需要 $O(n\ \log_2 n)$次算术运算等.

贪婪法(greedy algorithm) 寻求一些问题的最优解或可能解的一种直接算法.求最优解的问题,通常都有若干个输入和一些约束条件.任何满足这些约束条件的一个子集称为一个可能解,使目标函

数达到最大值或最小值的可能解称为最优解(这里的目标函数是问题中给定的).

贪婪法是一种多步决策方法,其每一步选择都应使得能构成问题的一个可能解,同时应使目标函数的值增加最多(当求目标函数最大值时)或增加最少(当求目标函数最小值时).这种选择过程是以某些最优化量度为根据的.这里使用的最优化量度,有时可以是目标函数本身,有时可以是别的量度.

最优化量度的选择是贪婪法的关键.一般地,贪婪法不能保证得到最优解,因为它为了提高效率,更快地找到解答,而没有充分考虑到各种复杂情形.对于某些特定输入,贪婪法可能给出最优解或接近最优的解,而对另一些特定输入,获得的解可能并不理想.贪婪法可用于求解背包问题、多处理机的调度、带时限的作业调度、磁盘文件的最佳存储等问题.

回溯法(backtracking) 算法设计的基本方法之一.采用系统地搜索给定问题的解空间的方法来确定问题的解.使用一种所谓解空间的树形结构将使这种搜索容易实现.搜索之前,先把各种候选对象组织成一棵树,每个树叶对应着一个候选对象,每个内部结点表示若干个候选对象(即在此结点下面的各个树叶所对应的候选对象).

回溯法是从树根开始按深度优先搜索的原则向下搜索,搜索的情况可概括为一句话"向前走,碰壁回头",即沿着一个方向尽量向下搜索,直到发现此方向上不可能存在解答时,就回到上一个结点,沿着另一个方向进行同样的工作.有许多问题,需要找出它的解集或者要求回答什么样的解是满足某些约束条件的最优解时,往往要使用回溯法来求解.一般地,回溯法效率较低,但是描述比较简明.这种方法适用于解的组合数相当大但仍然有限的那一类问题,例如八皇后问题、图的着色问题、哈密顿回路问题等.调试一些软件系统也采用回溯法.

分支定界法(branch and bound method) 用于解决组合最优问题的一种方法.它是一种构造性的搜索法,可在整个允许解空间中进行最优搜索.分支定界法与回溯法类似的是:都从所谓解空间的树形结构的树根开始向下搜索.不同的是:分支定界法比回溯法要复杂些,它进行最优搜索的原则是对解集合反复进行分支,每次分支时都对所得的子集合计算最优解的界.如果对某个子集合求得的界不优于已知的允许解,则对此子集合就不再进行分支;否则,继续分支以搜索更好的解,直到所得的子集合仅含有一个解时为止.为了提高效率,分支定界法常常利用一个适当选取的评估函数为最优解估界.评估函数选得好,就会很快地找到解答,选得不好,就可能找不到解答或者找到的不是最优解(有时它可以作为最优解的一个近似解).使用分支定界法时,要求实际问题具有组合性,解集合有可分支性,且对每个子集合能估计出最优解的界.

递归(recursion) 计算机问题求解的一种算法.是一种处理过程,这种过程的某一步要用到它自身的上一步(或上几步)的结果.一个直接或间接地调用自己的过程称为递归过程.在本过程中出现调用本过程自身的过程称为直接递归过程;若过程 P 包含着对另一过程 Q 的引用,而 Q 又直接或间接地引用 P,则称 P 为间接递归过程.一个使用函数自身给出定义的函数称为递归函数.在数学定义和计算机算法中,递归技术是一种特别有力的工具.递归函数或递归过程的应用往往使函数的定义和算法的描述比使用非递归方法更简明.例如,阶乘函数 $n!$ 的递归定义形式为

$$\begin{cases} 0! = 1; \\ n! = n \cdot (n-1)! \quad (n > 0, n \in \mathbf{Z}). \end{cases}$$

显然,递归的能力在于有可能用有限的语句来定义对象的无限集合.同样地,使用有穷的递归程序可以描述无穷多次计算.

递归过程(recursive procedure) 见"递归".

分类(sorting) 亦称整序或排序.是按规定的次序重新安排给定的一组对象的过程.分类的目的是便于随后对已分类的集合的成员进行检索.通常每个对象都包含一个称为键的标识符,它是某个线性有序集合中的元素,可以通过比较两个键值以确定哪一个更大或者二者是否相等.按要处理的数据是全部放在内存,还是放在外存,分类方法通常被分为内分类和外分类两大类.对于一个分类方法,如果在分类过程中,具有相等键的项的相对次序不被改变,则称分类是稳定的.如果项已被按某个次级键排序(分类)了,即已按与第一键本身无关的性质排序了,则常希望分类是稳定的.

整序(sorting) 见"分类".

内分类(internal sorting) 一种分类方法.是在整个分类期间,全部数据都存放在内存的分类方法.表示内分类方法的特性主要有时间复杂性和空间复杂性.后者要求节省可使用的存储空间(如采用就地分类技术);前者要求有较高的效率.对于效率的一个很好的度量是计算所需的键的比较次数 C 和项的移动(调换)次数 M.这两个数是被分类的项数 n 的函数.一些被称为直接方法的分类技术,一般都需要数量级为 n^2 次的键的比较;而好的分类方法仅需数量级为 $n\log_2 n$ 次比较($n\log_2 n$ 是分类法所需的比较次数的下界).现有的内分类方法很多,其中大多数已组成软件包(或程序库),下面仅列出几种重要而常用的分类技术的主要特性及适用条件,以供应用选择时参考:

1. 分布计数.它是稳定的,但要求为计数器和

$2n$ 个记录提供所需的存储空间. 当键的变化范围较小时，是非常有效的.

2. 直接插入. 编程简单，不需要额外的空间，对于较小的 n(例如 $n\leqslant 25$)十分有效.

3. 递减增量的分类. 亦称希尔分类. 容易编程，使用极小的空间，对于适当大的 n(例如 $n\leqslant 1000$)相当有效.

4. 表插入. 适用于小的 n，它通过处理链接减少了移动长记录的耗费，当记录有可变长度或者是其他数据结构的一部分时，特别有利.

5. 划分交换. 即霍尔方法，一般称为快速分类. 被认为是内分类中最有用的通用技术，它需要非常少的内存空间，而且在大多数情况下平均运行时间最少；但在最坏情况下运行得相当慢，所以每当可能是非随机的数据时，应当小心地选择划分的元素.

6. 堆分类. 需要极小的内存，运行的平均时间和极大时间都大体上是快速分类平均运行时间的两倍.

7. 基数分类. 特别适合于某种形式键的一种表示类，这些键或者比较短，或者是按一个非字典顺序整理的序列，或者也可以使用分布计数方法以代替链接；需要 $2n$ 个记录空间和一个计数器表. 适用于大的 n.

8. 直接选择. 当有特殊硬件可用来高速寻找一个表的最小元素时，是一个特别合适的简单方法.

9. 表合并. 和堆分类相似的一种表分类，能保证甚至在最坏的情况下也是相当快的，而且对于相等的键是稳定的.

10. 冒泡分类. 在文件几乎已分类的情况下，这种方法相当快.

11. 归并分类. 采用了分治策略，使得在最坏情况下复杂性为$O(n\ \log_2 n)$.

外分类(external sorting)　一种分类方法. 是对外存文件中大量数据的整序过程. 当输入文件的数据量超过了计算机内存容量时，不可能也不允许全部驻留在内存中，而必须将其存放在外存储器上，对此要采用外分类方法. 外分类期间，需要数据在内、外存之间移动. 在外分类中，最常用的分类方法是归并分类法. 这种方法通常经历两个阶段. 第一阶段是把文件逐段地输入到内存，用较好的内分类方法对输入文件段进行分类；整个文件经过逐段分类后(称为归并段)，又一段一段地写回到某个外存设备上. 第二阶段是对外存上的全部初始归并段使用某种归并方法进行多遍归并，最后在外存上形成整个文件的单一归并段，这样就完成了这个文件的外分类.

算法分析(analysis of algorithm)　计算机科学研究的重要内容之一. 是研究算法所耗费的计算资源与问题规模之间的函数关系. 耗费的计算资源指算法执行时所需要的时间和存储空间. 问题的规模，也称问题的大小，通常用一个自然数表示，作为问题输入数据量多少的度量. 算法分析的中心问题是研究当问题规模趋于无穷时，时间和空间的复杂度(参见“计算复杂性”)以何种速度增长. 复杂度的量级即极限属性，称为渐近复杂度.

算法分析的研究领域分为对某个具体算法的分析和对某类问题的算法分析两大类. 分析具体算法时，有两个常用的指标：一个是算法的平均行为；一个是算法的最坏行为. 这些分析往往是在假定输入数据满足某种概率分布的前提下进行的，也可根据输入数据的概率分布来求整个复杂度的分布. 至于对一类问题的算法分析主要是求复杂度的下界. 因为为这类问题找到的任何一个新的有效算法都可能提供一个新上界，而下界意味着任何算法都不会比它更好. 这是困难之处和关键所在. 分析算法的复杂度与计算模型有关，大多属于随机存取机器模型，这是一种确定型的串行计算模型.

渐近复杂度(asymptotic complexity)　见“算法分析”.

计算复杂性理论(computational complexity theory)　使用数学方法研究各类问题的计算复杂性的学科. 它对计算中所需的各种资源(计算时间、存储空间等)的耗费作定量的分析，并研究各类问题之间在计算复杂程度上的相互关系和基本性质，是算法分析的理论基础. 在计算一类问题时，资源耗费的多少与被计算问题本身的大小有关，它是问题大小的函数，称为问题对该资源需求的复杂度. 计算复杂性理论研究的主要内容包括对复杂度函数增长的阶进行分析，探讨它们对于不同的计算模型在一定意义下的无关性，根据复杂度的阶对被计算问题分类，研究各种不同资源耗费之间的关系，对一些基本问题的资源耗费情况的上、下界作估计等.

计算复杂性(computational complexity)　计算机科学理论中的一个重要概念. 它是各种计算中所需的各种计算资源的耗费量的估计，表示计算问题的计算难度. 计算资源耗费的多少可表示成计算问题大小 n 的函数，称为问题对该资源需求的复杂度，包括时间复杂度和空间复杂度. 常用的时间和空间的度量方式是均匀耗费标准：执行一条指令算作耗费一个单位的时间，使用一个内存单元算作耗费一个单位的空间；另一种度量方式是对数耗费标准. 复杂度函数的计算方式有两种：

1. 最坏情况复杂度，是规模为 n 的所有问题的复杂度的最大值.

2. 平均情况复杂度，是规模为 n 的所有问题的复杂度的平均值.

当规模为 n 增加时，复杂度的量级即极限属性，称为渐近复杂度.通常认为，多项式复杂度的算法是现实可行的，而指数复杂度的算法是现实不可行的.

计算复杂性的研究结果可直接用于算法分析和优化.这方面的研究开始于20世纪60年代初期，开始时侧重于研究一般的可计算函数的计算复杂性，这项研究构成了可计算函数论的一部分.20世纪70年代开始了机器计算复杂性的研究.这既包括对计算方法中各类算法的分析和优化，也包括对计算机硬件和软件设计中所用算法的分析和优化.目前重点是关于并行计算的复杂性的研究.这一类研究有较大的实际意义，已发展成为计算机科学的一个领域.

复杂性度量(complexity measure) 计算复杂性的衡量标准(参见"算法分析"、"计算复杂性理论"、"计算复杂性"等).这种衡量标准不能表示为绝对的数量大小，而应表示为问题大小 n 的一个函数.例如对一个问题的某个算法所消耗的时间的度量，不应依赖于计算工具的计算速度，而应将算法的主要运算次数表示为 n 的函数.在用普通的算法计算两个 n 阶方阵的乘积的过程中，其运算次数约为

$$2n^3+2n^2+n,$$

它是 n 的三阶多项式，忽略低次项后，可写为$O(n^3)$.这里 $O(n^3)$表示是与 n^3 同阶的一个量，O 可以认为是与 n 无关的某个常量.

时间复杂性(time complexity) 计算机算法分析的重要概念之一.是计算中所需的计算时间资源耗费量的估计.如果一个问题的大小为 n，解这个问题的某一算法所需时间为 n 的某个函数$T(n)$，则称 $T(n)$为该算法的时间复杂性(参见"计算复杂性"、"复杂性度量"、"内分类"等).例如，n 个整数的分类问题，当用归并分类算法时，时间复杂性为 $O(n\log_2 n)$，当用直接插入算法时，时间复杂性为 $O(n^2)$.

空间复杂性(space complexity) 计算机算法分析的重要概念之一.是计算中所需的存储空间资源耗费量的估计.如果一个问题的大小为 n，解这个问题的某一算法所需的辅助存储空间为 n 的某个函数$S(n)$，则称 $S(n)$为该算法的空间复杂性(参见"计算复杂性"、"复杂性度量"、"内分类"等).例如，n 个整数的分类问题，当用归并分类算法时，空间复杂性为 $O(n)$，当用直接插入分类算法时，空间复杂性为 $O(1)$(一个与 n 无关的常数).

问题复杂性(problem complexity) 计算机问题求解的重要概念之一.是计算一个问题的所有算法中，时间复杂性最小的那个算法的复杂性(参见"计算复杂性"、"复杂性度量"、"时间复杂性"等).例如，在 n 个任意的整数中找出最大的数和最小的数，$[3n/2-2]$次比较运算是必须的，因此这个问题的复杂性是$O(n)$.又如著名的梵塔问题，2^n-1 次移动盘片是必须的，因此梵塔问题的复杂性是 $O(2^n)$.

计算资源(resource on the computation) 计算复杂性理论的一个术语.指计算中所需的各种资源.一般地，各种计算模型的主要资源有并行时间、串行时间和空间三种.

1. 并行时间和巡回.并行时间一般指并行模型计算时所需步数，例如，向量机的自始至终执行指令的总条数.但对串行模型也可以定义一种称为巡回的资源.可以证明它相当于并行时间.对于多带图灵机，它是工作带头部改变方向的次数.一般地，巡回是周相的总数，而周相则是串行模型工作中的一个阶段，在此阶段中计算出来而记录在工作空间上的信息，不再被读到.

2. 串行时间.计算过程中原始运算的总量.对于串行模型而言，它代表计算自始至终的总步数；对于并行模型而言，每一步可以同时作许多个原始的运算，自始至终各步的原始运算数目的总和就是串行时间.

3. 空间.在计算过程中需要记录下来以备后用的最大中间信息量.对于多带图灵机，是计算过程中用过的工作带上的方格数.

NP 完全性(NP-completeness) 计算复杂性理论中的一个重要概念，它表征某些问题的固有复杂度.由于"$NP\overset{?}{=}P$ 问题"难以解决，人们从 NP 类的问题中分出复杂性最高的一个子类，称为 NP 完全类.已经证明，任取 NP 类中的一个问题，再任取 NP 完全类中的一个问题，则一定存在一个确定型图灵机上的具有多项式复杂性的算法，把前者转换成后者.最早提出 NP 完全性问题的是库克(Cook, S. A.)，他在1971年证明了命题演算的可满足性问题具有 NP 完全性.

NP 完全性的研究，在理论和实践两个方面都具有重要意义.理论方面，已经证明，只要有一个 NP 完全问题属于 P，就能推出 $NP=P$；反之，要证明 $NP\neq P$，一个明显的方法，就是在 NP 中找一个不属于 P 的问题.实践方面，在算法设计和分析过程中，如果已证明某问题是 NP 完全的，这就意味着面临的是一个具有相当复杂程度的困难问题，对于它，要找出一个在计算机上可行的(即多项式时间界的)算法是十分困难的，甚至可能根本找不到.因此，对于 NP 完全问题，最好是寻找近似解法，或者针对该问题的某些有实用价值的特殊情况，寻找多项式时间算法.

NP 完全问题(NP-complete problem) NP 中最难问题的一种形式化.对于一个问题 q，如果 q 属

于 NP，且 NP 中任意一个问题都能够用多项式时间归约到 q，则称 q 为 NP 完全的，或 q 具有 NP 完全性. 在研究著名的"$NP \stackrel{?}{=} P$ 问题"中，许多人猜测 $NP \neq P$，即在 NP 中有不是多项式时间可解的问题. 在直觉上如果这种问题存在的话，它就是 NP 中"最难的"问题.

历史上第一个 NP 完全问题是由库克(Cook，S. A.)于 1971 年提出的可满足性问题. 有代表性的 NP 完全问题还有旅行商问题，带优先次序的调度问题，图的着色、装箱以及背包问题等. 这类问题来自计算机科学、数学、逻辑学等许多学科领域，目前总数已达数千.

可计算性理论(computability theory) 亦称算法理论或能行性理论. 计算机科学的理论基础之一. 是研究计算的一般性质的数学理论. 可计算性理论通过建立计算的数学模型，精确地回答什么是计算、哪些是可计算问题、哪些是不可计算问题等. 计算的过程就是执行算法的过程. 因此，算法这一直观概念的精确化是可计算性理论的重要课题之一.

20 世纪 30 年代，为了讨论对于每个问题是否都有解决它的算法，数学家和逻辑学家们提出了几种不同的算法定义. 哥德尔(Gödel，K.)和克林(Kleene，S. C.)提出了递归函数的概念，丘奇(Church，A.)提出 λ 转换演算，图灵(Turing，A. M.)和波斯特(Post，E. L.)各自独立地提出了抽象计算机的概念，并且证明了这些数学模型的等价性. 后来，人们又提出了许多等价的数学模型，如马尔可夫(Марков，A. A.)于 20 世纪 40 年代提出的正规算法、20 世纪 60 年代前期提出的随机存取机器模型等. 把算法看成抽象计算机的程序，能在抽象计算机上编出程序计算其值的函数称为可计算函数.

可计算性理论研究的主要内容有递归函数、图灵机、波斯特系统和判定问题等. 可计算性理论不仅是建立计算机的理论模型，它也给程序及形式语言提供了理论模型，各种自动机及形式语言的语法就是在图灵机及波斯特系统的基础上发展起来的. 可计算性理论中的基本思想、概念和方法被广泛应用于计算机科学的各个领域. 递归的思想被用于程序设计，产生了递归过程和递归的数据结构，也影响了计算机的体系结构. 建立数学模型的方法在计算机科学中被广泛采用. λ 转换演算被用于研究程序设计语言的语义，例如，表处理语言就以 λ 转换演算为理论基础.

算法理论(algorithmic theory) 即"可计算性理论".

递归函数(recursive function) 可计算性理论中的一个概念(参见《数学辞海》第四卷同名条).

1. 一种计算模型. 是处处有定义的部分递归函数. 根据丘奇论题，递归函数就是可计算函数. 已经证明，作为计算模型，递归函数与图灵机及波斯特系统都是等价的.

2. 具有递归性质的函数. 即每次求函数值时，要用到前一次或前几次的迭代结果.

原始递归函数(primitive recursive function) 可计算性理论中的概念之一. 是由一组初始函数及由此通过有限次的复合运算和原始递归运算得到的函数. 其中：

1. 初始函数是零函数，即 $O(x)=0$，后继函数 $S(x)=x+1$，投影函数 $U_j^{(n)}(x_1,x_2,\cdots,x_n)=x_j$.

2. 代入运算. 已知 g_i,h 为原始递归函数($1\leqslant i\leqslant m$)，则通过复合运算得到的 $f(x_1,x_2,\cdots,x_n)=h(g_1(x_1,x_2,\cdots,x_n),\cdots,g_m(x_1,x_2,\cdots,x_n))$ 也是原始递归函数.

3. 原始递归运算. 已知 g,h 为原始递归函数，则通过原始递归运算得到的函数

$$\begin{cases}f(x_1,x_2,\cdots,x_{n-1},0)=g(x_1,x_2,\cdots,x_{n-1}),\\ f(x_1,x_2,\cdots,x_{n-1},x_n+1)\\ =h(x_1,x_2,\cdots,x_{n-1},x_n,f(x_1,x_2,\cdots,x_n))\end{cases}$$

也是原始递归函数.

原始递归函数是数论函数的一部分. 自变量值和函数值都是自然数的函数，称为数论函数. 许多常用的处处有定义的函数都是原始递归函数，但并非一切直观可计算的、处处有定义的函数都是原始递归函数.

判定问题(decision problem) 可计算性理论中的一个研究内容. 对于某个问题，研究是否存在解决的办法. 判定问题是针对一个谓词来讲的. 若 $P(x_1,x_2,\cdots,x_n)$ 是递归谓词，就称与谓词 $P(x_1,x_2,\cdots,x_n)$ 相关的判定问题是(递归)可解的或(递归)可判定的；否则就称这个判定问题是(递归)不可解的或(递归)不可判定的. 例如，令 $P(x)$ 代表谓词"x 是素数"，相应的判定问题是：有没有算法可以判断任意 x 是否是素数？由于 $P(x)$ 是递归谓词，所以这个判定问题是递归可解的. 判定问题也可以针对集合来讲：若 S 是递归集，就称与这个集合相关的判定问题是(递归)可解的或(递归)可判定的；否则，就称这个判定问题是(递归)不可解的或(递归)不可判定的. 集合与谓词有密切的联系，"$x\in S$"就是一个谓词. 与集合 S 相关的判定问题实际上就是与这个谓词相关的判定问题. 因此，研究一个判定问题，也就是研究一个谓词是否是递归谓词；或研究一个集合是否是递归集；或者研究一个函数是否是递归函数. 大多数判定问题是不可判定的. 证明不可判定，也就是证明一个谓词、或一个集合、或一个函数不是递归

的.主要的证明方法,除采用较困难的直接法外,更多的是采用间接方法.所谓间接方法就是利用一个判定问题的不可解,证明另一个判定问题不可解,通常利用归纳方法.

自20世纪70年代末以来,判定问题已成为计算机科学的一个重要的组成部分.计算机科学中提出的大量的判定问题,尽管其答案往往是否定的,但仍有重大的实际意义.例如,任意两个程序是否等价的问题,这是一个不可判定的问题.也就是说不存在有效的方法可以用来判断任何两个程序是否等价.既然如此,就不要花力气去寻找这样的方法了.

停机问题(halting problem) 可计算性理论中的一类判定问题.停机问题有多种不同的叙述方式和证明方法,它们分别适用于具有不同特征的问题.例如,当把带和图灵机连结起来,或者程序在通用数字计算机中运行时,必须要回答的问题是机器将会停机或永不休止地运行.要做的判定是给定任何机器和一条原始带表达式,机器将停机或不停机.这就是停机问题.又如,是否存在一个算法,对于任意给定的图灵机都能判定任意的初始格局是否导致停机,这个问题就是图灵机的停机问题.图灵(Turing, A. M.)于1936年证明,图灵机的停机问题是不可判定的.停机问题是研究许多不可判定问题的基础,由它可以推出计算机科学、数学、逻辑学中的许多问题是不可判定的.人们往往把一个问题的判定归结为停机问题:"如果问题 A 可判定,则停机问题可判定".从而证明问题 A 的不可判定性.

自动机理论(automaton theory) 关于自动机的功能、结构及两者关系的数学理论.自动机是一个数学概念,它是离散数字系统的抽象模型.这里所说的离散数字系统,乃是一种动态系统,它的变量是数字量,时间是离散的.例如,数字电路和算法就是两个典型的离散数字系统.自动机理论的主要研究课题是分析和综合问题:给出一个具体的自动机的结构,分析它的功能;给出自动机的功能描述,综合出能实现此功能的自动机的结构.

自动机理论是理论计算机科学中较早形成的部分.早在1850年,英国布尔(Boole, G.)就在用数学方法研究思维规律的问题时,建立了逻辑代数.1948年,美籍匈牙利数学家冯·诺伊曼(von Neumann, J.)提出建立自动机的一般逻辑理论.20世纪50年代,在开关网络理论和数理逻辑中图灵机理论的基础上,形成了自动机理论这一数学分支学科.20世纪50年代以来,自动机理论有了深入的发展和广泛的应用.自动机理论大致可分为以下五个次级学科:

1.有限自动机理论.主要研究对象为开关网络、数字电路、计算机这类存储量有限的自动机.

2.无限自动机理论.主要研究对象为算法和理想计算机这类存储量不受限制的自动机.

3.概率自动机理论.主要研究对象是在环境或内部具有随机因素的自动机.

4.细胞自动机理论.主要研究对象是由许多互连的小自动机并行运算形成的大自动机.

5.抽象自动机理论.将自动机作为一种数学系统,研究自动机的一般数学性质.

自动机理论与数理逻辑、可计算性理论、计算复杂性理论、形式语言理论、控制论等数学分支都有关系,特别是它与形式语言理论关系密切.一方面自动机作为形式语言的一种主要描述方法,另一方面形式文法也可作为自动机识别集的一种描述方法.自动机理论在自动控制、计算机和数字通信等领域有着广泛的应用.

有限自动机(finite automaton) 亦称时序机.有限离散数字系统的抽象数学模型.一个有限自动机 M 由五元组(X,Y,S,δ,λ)给定,其中 X,Y 和 S 都是非空有限集,分别称为 M 的输入集、输出集和状态集;δ 是笛卡儿积集合 $S\times X$ 到 S 的映射,称为 M 的下一状态函数;λ 是 $S\times X$ 到 Y 的单值映射,称为 M 的输出函数.当 δ 是单值映射时,称 M 为确定型有限自动机;当 δ 是多值映射时,称 M 为非确定型有限自动机.有限自动机有三种功能:作为序列转换器,将输入序列变换为输出序列;作为序列识别器,识别输入的序列是否具有某种性质;作为序列产生器,产生具有所要求性质的序列.

研究有限自动机的功能、结构以及两者关系的数学理论称为有限自动机理论.有限自动机理论的基本内容包括逻辑网络、状态化简、状态分配、神经网络和有限识别器等.

1.逻辑网络.基本的逻辑元件按是否具有记忆功能,可以分为记忆元件(如触发器和延迟器等)和组合元件(如各种与、或、非门等)两类.把一些基本逻辑元件按一定要求连结起来,就组成逻辑网络.若把逻辑网络中进入记忆元件的输入线去掉后所得网络不再含有回路,则称这样的网络为合式网络.不含记忆元件的合式网络称组合网络.逻辑网络比组合网络复杂.在工程实现上,要求对于一个给定的有限自动机建立和实现此有限自动机的逻辑网络.已经证明任何合式网络的功能都可以用一个有限自动机来描述;任何一个有限自动机描述的功能也都可以用合式网络来实现.

2.状态化简.对任何有限自动机都惟一(在同构意义下)存在一个状态数目最少的有限自动机与它等价.根据有限自动机理论,对给定的有限自动机,可有效地求出与之等价的最简形式的有限自动机.

3.状态分配.要构造具有多个状态的网络,需要使用多个基本记忆元件,利用这些记忆元件的各种

状态组合来表示不同的状态. 一般地，不同的状态分配导致逻辑网络具有不同的复杂程度. 如何选择较好的分配方案，使逻辑网络的构造尽可能地简单，是有限自动机研究的一个主要课题.

4. 神经网络. 1943 年，麦克卡洛克(Mcculloch)和皮特斯(Pitts，W.)提出的神经网络模型是有限自动机的一个实例. 1951 年，克林(Kleene，S. C.)在这种神经网络模型的基础上，提出了正则事件(正则语法)的概念，证明了正则事件是可以被神经网络或有限自动机表示的事件，而且神经网络或有限自动机可以表示的事件也一定是正则事件.

5. 有限识别器. 在形式语言理论中，有限自动机通常作为语言的识别器来使用. 作为识别器，有限自动机的输出可以被忽略，而由最后达到的状态去决定输入序列是否具有给定的性质，这种有限自动机也称为有限接收机. 按其下步状态是否完全确定，有限识别器可分为确定型和非确定型两种，它们分别与确定型和非确定型有限自动机相对应，它们也都接受同一类语言，即正则语言.

时序机(time series machine)　即“有限自动机”.

概率自动机(probabilistic automaton)　亦称随机自动机. 一种自动机. 是所处环境或内部具有有限或无限的随机因素的自动机，与非概率型自动机的主要区别是：概率自动机的动作是随机的. 每个概率自动机一般都需规定两组概率：一是给定自动机的初始状态的概率分布——初始分布，一般用一个随机矢量表示；二是规定在自动机处于某一状态，并向自动机输入某个字母的条件下，自动机下一动作(如状态转移、输出某个字母、改写字母等)的条件概率函数. 有了这两组概率，就可计算自动机到达某个最终状态的概率. 包含有不可靠元件的数字电路和通信的信道都可以表示为概率自动机.

随机自动机(random automaton)　即“概率自动机”.

细胞自动机(cellular automaton)　一种自动机. 是有限细胞空间形式的自动机. 常作为并行计算机的一种理论模型. 细胞空间概念是冯·诺伊曼(von Neumann，J.)在 20 世纪 50 年代初期研究自繁殖自动机的逻辑问题时提出的. 细胞自动机由许多细胞(或单元)构成. 每个“细胞”是一台计算机的模型，都有自己的存储器，都具有输入、加工和输出数据的能力. 细胞和细胞之间有邻接关系，有些模型还和外部相连，以便与外部交换输入输出数据. 冯·诺伊曼细胞自动机是最早最基本的自动机，其他各种类型的细胞自动机(如 L 系统)都是由冯·诺伊曼自动机推广而来的. 细胞自动机除了在形式上可作为并行计算机的理论模型来研究外，还可以作为语言(被机器接受的输入字的集合)识别器，用于模式识别. 此外，它对于大规模集成电路的设计方法也具有重要的意义.

下推自动机(pushdown automaton)　简称 PDA. 亦称后进先出自动机. 一种自动机. 是能控制一条输入带和一个栈的有限自动机. 栈也称为“后进先出”表，即符号的写入或取出只能在表的顶端进行. 在文法结构数学模型中，下推自动机可以用作上下文无关语言的识别接受器. 下推自动机可以形式地定义为一个 7 元组 $P=(Q,\Sigma,\Gamma,\delta,q_0,Z_0,F)$，其中：

1. Q 是一个有限的状态集合.

2. Σ 是一个有限的输入字母表.

3. Γ 是一个有限的栈字母表.

4. δ 是一个从 $Q\times(\Sigma\cup\{e\})\times\Gamma$ 到 $Q\times\Gamma^*$ 的有限子集的映射.

5. q_0 在 Q 中，是有限控制的初始状态.

6. Z_0 在 Γ 中，是一个称为栈底的符号.

7. $F\subseteq Q$，是终结状态集合.

下推自动机是一个非确定的装置. 对于有限自动机，确定型和非确定型有限自动机在接受语言上是等价的. 这一结论对于下推自动机是不成立的. 非确定型下推自动机接受的语言集合是上下文无关语言；确定型下推自动机接受的语言集合是非确定型下推自动机接受的语言集合的真子集，称为确定型上下文无关语言.

后进先出自动机(last in first out automaton)即“下推自动机”.

线性有界自动机(linear bounded automaton)一种图灵机(参见“图灵机”). 是把计算限制在仅仅包含输入的那一段带上的图灵机. 可用作上下文有关语言的识别接受器. 线性有界自动机(缩写为 LBA)可形式地由 $M=(K,\Sigma,\Gamma,\delta,q_0,F)$来表示. 其中：$K$ 是状态的有限集；Γ 是带符号的有限集；$\Sigma\subseteq\Gamma$ 是输入符号集；K 中的 q_0 是起始状态；$F\subseteq K$ 是终结状态集；δ 是从 $K\times\Gamma$ 到 $K\times\Gamma\times\{L,R\}$ 子集的映射，(L,R)分别是读写头左右移一格. Σ 含有两个特殊的符号，通常记为 ₵和 \$，它们分别是左端标志和右端标志. 这些符号开始就处在输入带的端点，其作用是阻止带头离开带上出现符号的区.

确定型和非确定型图灵机接受的语言是相同的. 但是，已知非确定型线性有界自动机(以 NDLBA 表示)接受的语言类正好是上下文有关语言. 确定型线性有界自动机(以 DLBA 表示)的功能不会超过 NDLBA 的，以不等式表示为

$$L(\text{DLBA})\subseteq L(\text{NDLBA}),$$

其中 L 表示该类自动机接受的语言集合. 至今人们仍未能把包含关系精确为真包含关系或相等关系.

这是一个著名的尚未解决的问题,简称 LBA 问题.

图灵机(Turing machine) 一种抽象计算模型. 是 20 世纪 30 年代中期英国数学家图灵(Turing, A. M.)首先提出来的,用来精确定义可计算函数. 图灵机由一个控制器、一条可无限延伸的存储带和一个读写头组成. 它所能进行的操作为:

1. 左移(读写头在存储带上向左移一格).
2. 右移.
3. 在存储带的某一格内写下或清除一个符号.
4. 条件转移等.

图灵机的结构虽然比较简单,但在理论上它却能够模拟现代数字计算机的一切运算,因此可看做是现代数字计算机的一种数学模型. 可以通过对这种模型的研究来揭示数字计算机的性质. 可以证明,存在一个图灵机 U,它可以模拟任何其他图灵机,这样的图灵机 U 称为通用图灵机. 通用图灵机正是后来出现的存储指令的通用数字计算机的理论原型. 图灵机所定义的语言类称为递归可枚举集. 图灵机所计算的整数函数类称为部分递归函数. 图灵机在计算能力上等价于递归函数、波斯特系统等其他计算模型.

波斯特-图灵程序(Post-Turing program) 一种用于字符串操作的抽象机的程序. 假设这种抽象机有一条双向无穷带作为存储装置,带上划分为若干小格,每个格子上可以写一个符号,这种抽象机还有一个带头,每一时刻它都指向一个小格,并可以左右移动. 抽象机仅有四条指令,它们的功能如下:

PRINT σ:在带头指向的格子上写符号 σ.

IF σ GoTo L:如果带头指向的格子上为符号 σ,则转到标号为 L 的指令去执行;否则,执行下一条指令.

RIGHT:带头向右移一小格.

LEFT:带头向左移一小格.

开始时,带上可以放若干个符号作为输入,并可以指定带头指向的符号. 用以上四种指令(加上标号),可以编出各种波斯特-图灵程序. 执行程序后,带上的非空白符号就是程序的输出.

形式语言理论(formal language theory) 计算机科学的内容之一. 是用数学思想和数学方法来研究自然语言和人工语言的语法的理论. 形式语言理论来源于数理语言学中的乔姆斯基理论. 在形式语言理论中,语言被看成是一个抽象的数学系统,乔姆斯基(Chomsky, N.)把语言定义为:按一定规律构成的句子或符号串的有限的或无限的集合. 形式语言理论就是对符号串的集合、它们的表示法、结构以及特性的研究.

文法和自动机是形式语言理论的基本内容. 文法用于生成语言,而自动机则用于识别语言. 文法是有限变换规则的集合,变换规则递归地生成潜在的无限的句子,并排除语言中的不成立句子. 乔姆斯基首先提出把文法和语言分为 0 型、1 型、2 型、3 型等四种类型. 后来,人们发现乔姆斯基分类和自动机的分类有密切的关系,并且证明了文法的生成能力和语言自动机的识别能力的等价性的四个重要结果:

1. 若一语言能为图灵机识别,则它就能由 0 型文法生成,反之亦然.

2. 若一语言能为线性有界自动机识别,则它就能由 1 型文法生成,反之亦然.

3. 若一语言能为下推自动机识别,则它就能由 2 型文法生成,反之亦然.

4. 若一语言能为有限自动机识别,则它就能由 3 型文法生成,反之亦然.

形式语言理论对计算机科学有重要的意义. 特别是在人们发现算法语言 ALGOL 60 中使用的巴科斯-瑙尔范式恰好与形式语言理论中的上下文无关文法等价之后,形式语言理论在计算机科学界开始广泛地应用. 形式语言理论已发展成为理论计算机科学的一个重要分支. 它在自然语言理解与翻译、程序设计语言的设计与编译、算法分析、图象识别、人工智能、社会和自然现象的模拟等方面有广泛的应用.

形式语言(formal language) 人工语言的总称. 是在一个有限的字母表上,按照一定的规则定义的人工语言的总称(参见“形式语言理论”、“乔姆斯基分类”和“文法”等). 形式语言定义的方法有许多种,例如,由图灵机识别或由 0 型文法生成的语言称为递归可枚举集或短语结构语言,由有限自动机识别或由 3 型文法生成的语言称为正则语言等. 另外,正则语言也可由正则表达式表示,上下文无关语言也可由 2 型表达式表示.

乔姆斯基分类(Chomsky classification) 亦称乔姆斯基层次. 一种分层方法. 是乔姆斯基(Chomsky, N.)于 1959 年提出的关于形式语言及其文法的分层方法. 乔姆斯基把文法定义为四元组 $G=(\Sigma, V, S, P)$,其中 Σ 是终结符号集,V 是非终结符号集,S 是开始符号,P 是变换规则(又称产生式)集. 这里的 Σ, V, P 都是有限集,$\Sigma \cap V=\varnothing$($\varnothing$ 表示空集合),$S\in V$. 若令 $\alpha\in(\Sigma\cup V)^{+}$, $\beta\in(\Sigma\cup V)^{*}$(用+代替 * 表示不含空字),则 P 中所有的产生式皆形如 $\alpha\rightarrow\beta$,表示 α 由 β 组成或用 β 替换 α. 这样定义的文法称为乔姆斯基文法,有时也称为短语结构文法或无限制文法. 对文法产生式作某些限制,可以得到生成能力各不相同的几类文法. 乔姆斯基按文法产生式的形式把文法分成以下四种类型:

1. 0 型文法. 即无限制文法.

2. 1 型文法. 限制其所有产生式均取 $\gamma A\delta\rightarrow\gamma\omega\delta$

的形式，其中 $\gamma,\omega,\delta\in(\Sigma\cup V)^*$，$|\omega|>0$（$|\omega|$表示符号串 ω 的长度），$A\in V$.

3. 2 型文法. 限制 1 型文法中产生式均取 $A\to\omega$ 的形式.

4. 3 型文法. 限制 2 型文法中产生式的右端为 aB 或 a，其中 $a\in\Sigma,B\in V$.

乔姆斯基把由 0 型文法、1 型文法、2 型文法和 3 型文法所生成的语言，分别称为 0 型语言、1 型语言、2 型语言和 3 型语言. 由于从 1 型到 3 型的限制条件是逐渐增加的，因此，不论对于文法或对于语言来说，都存在着如下的包含关系：

$$0\text{ 型}\supseteq 1\text{ 型}\supseteq 2\text{ 型}\supseteq 3\text{ 型}.$$

上述四种类型的文法及其所生成的语言的卓越见解，是乔姆斯基对于形式语言理论的重要贡献.

乔姆斯基层次(Chomsky hierarchy)　即“乔姆斯基分类”.

文法(grammar)　亦称形式文法. 定义形式语言（区别于自然语言）的一组规则. 在一个有限的字母表上，它能构成的字符串的个数是无限的. 为了能在这无限多的字符串中确定哪些字符串属于人们要定义的语言（一般也是无限多个字符串），需要用一种有限形式来刻画，这种有限形式（一组规则）就是文法. 最典型的文法就是乔姆斯基的文法谱系（参见“形式语言理论”和“乔姆斯基分类”等），它是由一个四元组 $G=(\Sigma,V,S,P)$ 来表示的. 其他的表示方法还有描述程序语言的巴科斯-瑙尔范式和语法图等.

形式文法(formal grammar)　即“文法”.

无限制文法(unrestricted grammar)　即“0 型文法”（参见“乔姆斯基分类”）.

短语结构文法(phrase structure grammar)　即“0 型文法”（参见“乔姆斯基分类”）.

上下文有关文法(context-sensitive grammar)　形式语言理论中一种形式文法. 用来描述上下文有关语言，在乔姆斯基分类中称为 1 型文法（参见“乔姆斯基分类”）. “上下文有关”这个术语来源于这种文法的产生式形式 $\gamma A\delta\to\gamma\omega\delta$，其直观意义是：只有在“上下文”$\gamma$-$\delta$ 中，才允许用 ω 去替换 A. 由于 ω 不能为空符号串，即 $|\omega|>0$，因此符号串 $\beta=\gamma\omega\delta$ 的长度至少等于或大于符号串 $\alpha=\gamma A\delta$ 的长度. 这就是说，上下文有关文法是由短语结构文法（0 型文法）的产生式加上限制 $|\beta|\geqslant|\alpha|$ 后所得到的文法. 上下文有关文法与上下文无关文法之间存在着如下关系：

1. 每一个上下文无关文法都包含于上下文有关文法之中.

2. 存在着不是上下文无关语言的上下文有关语言（例如，$L_1=\{a^nb^nc^n\mid n\geqslant1\}$ 是上下文有关语言而不是上下文无关语言）.

上下文无关文法(context-free grammar)　形式语言理论中一种重要的形式文法. 用来描述上下文无关语言，在乔姆斯基分类中称为 2 型文法（参见“乔姆斯基分类”）. 这种文法之所以是上下文无关的，是因为它的产生式皆形如 $A\to\omega$，其左端项是孤立的非终结符，可以被右端的符号串所替换，而不管左端项 A 所出现的上下文环境. 上下文无关文法可以化为以下两种简单的范式之一：其一是乔姆斯基范式，它的产生式均取 $A\to BC$ 或 $A\to\alpha$ 的形式；其二是格拉巴赫范式，它的产生式均取 $A\to aBC$ 或 $A\to\alpha$ 的形式. 其中，$A,B,C\in V$，是非终结符，$a\in\Sigma$ 是终结符；$\alpha\in\Sigma^*$ 是终结符串.

上下文无关文法被广泛地应用于描述自然语言和计算机语言. 利用乔姆斯基范式，可把任何上下文无关文法的推导树简化为二叉的推导树，从而有利于对自然语言进行层次分析. 由于程序设计语言的语法基本上都是上下文无关文法，因此上下文无关文法广泛地用于定义程序设计语言，使语法分析概念形式化，简化程序设计语言的翻译及其在字符串处理中的应用.

正则文法(regular grammar)　形式语言理论中一种形式文法. 用来描述正则语言，在乔姆斯基分类中称为 3 型文法（参见“乔姆斯基分类”）. 若限制产生式的形式为 $A\to aB$ 或 $A\to a$，其中 A,B 是非终结符，a 是终结符，那么文法为右线性文法. 若限制所有产生式的形式均为 $A\to Ba$ 或 $A\to a$，那么文法为左线性文法. 每一右线性文法必有与之等价的左线性文法存在，即是说，这两种文法生成相同的语言——正则语言. 右线性文法和左线性文法都称为正则文法.

正则表达式(regular expression)　用来表示正则语言的一种简单表达式. 正则语言是正则集. 设 Σ 是一个有限集，则 Σ 上的正则表达式及它所表示的正则集可递归地定义如下：

1. $\varnothing$是一个正则表达式，表示空正则集$\varnothing$.

2. e 是一个正则表达式，表示只含空元素的正则集$\{e\}$.

3. Σ 的元素 a 是一个正则表达式，表示正则集$\{a\}$.

4. 如果 r 和 s 分别是表示正则集 R 和 S 的正则表达式，那么$(r+s)$，$(r\cdot s)$和(r^*)就分别是表示正则集 $R\cup S$，RS 和 R^* 的正则表达式（运算符$\cup$，$\cdot$，$*$ 分别为并、连结、乘幂闭包. 式子里连结符 $\cdot$ 可以略去不写，运算的优先顺序为：$*$，$\cdot$，$\cup$）.

设 α,β 和 γ 都是正则表达式，下列等式描述了正则表达式的某些基本代数性质：

1. $\alpha+\beta=\beta+\alpha$.

2. $\varnothing^*=e$.

3. $\alpha+(\beta+\gamma)=(\alpha+\beta)+\gamma$.

4. $\alpha(\beta\gamma)=(\alpha\beta)\gamma$.

5. $\alpha(\beta+\gamma)=\alpha\beta+\alpha\gamma$.

6. $(\alpha+\beta)\gamma=\alpha\gamma+\beta\gamma$.

7. $\alpha e=e\alpha=\alpha$.

8. $\varnothing\alpha=\alpha\varnothing=\varnothing$.

9. $\alpha^{*}=\alpha+\alpha^{*}$.

10. $(\alpha^{*})^{*}=\alpha^{*}$.

11. $\alpha+\alpha=\alpha$.

12. $\alpha+\varnothing=\alpha$.

程序设计理论(theory of programming) 亦称程序理论.程序设计方法学的理论基础.其主要研究内容包括程序设计的方法、程序设计语言的理论以及程序正确性理论等.

1. 程序设计的方法.针对各类问题所具有的不同特点,概括出各类抽象模型,研究它们的特性,讨论适用于它们的程序设计的方法.

2. 程序设计语言的理论.包括对程序设计语言的语法和语义的研究,是设计、实施和使用程序设计语言的理论基础.

3. 程序正确性理论.研究编制正确程序的两种主要途径:程序验证和程序综合.前者研究如何使用数学推理来严格论证程序是否符合其目标;后者研究如何由给定目标出发,逐步构造出一个可在计算机上运行的正确无误的程序.

程序理论(program theory) 即“程序设计理论”.

程序设计方法学(programming methodology) 研究程序设计各阶段工作的原理、原则和技术的一门学科.其目标是用科学化和工程化的程序设计方法,设计出可靠易读且代价合理的程序.程序设计方法大致可分为两类:一类是全局性的,如结构程序设计方法,它不仅要求编出的程序结构良好,而且要求程序设计过程也是结构化的、层次式的、逐层降低抽象级别的;另一类则是局部性的,如子程式方法、协同程式方法、顺序程序设计、并发程序设计和分布式程序设计等.程序设计方法学不仅研究各种具体的方法,而且着重研究各种具体方法的共性、涉及规范的全局性方法,以及这些方法的现实背景和理论基础.它包括程序理论、研制技术、资源环境、工程规范和自动程序设计等课题.

程序逻辑(program logic) 亦称霍尔逻辑.描述和论证程序行为的逻辑.这种逻辑是在1969年前后,由英国霍尔(Hall,C. A. R.)首次给出的一类程序语言的逻辑系统的基础上发展起来的.20世纪70年代初,波兰和瑞士的一些学者第一次把模态逻辑引入程序逻辑,使用算法逻辑描述和分析程序.20世纪70年代中期,人们进一步提出使用动态逻辑和时态逻辑来描述和论证程序.程序逻辑与古典的谓词逻辑的区别在于,前者之中包含有时间概念(或者说“过程”概念或“程序”概念),而后者之中则没有.

程序逻辑的基本方法是先给出建立程序和逻辑间联系的形式化方法,然后建立程序逻辑系统,并在此系统中研究程序的各种性质.例如,在霍尔逻辑中,程序逻辑公式的形式为$\{P\}S\{Q\}$,其中P,Q代表命题,S代表程序.此公式表示如果在程序S执行前命题P成立,则在S执行完毕后,命题Q就成立.由于这类公式不能描述S的终止性,所以霍尔逻辑只是讨论程序部分正确性的逻辑.它实际上只描述了程序S的规范,并没有彻底把程序和逻辑统一起来.改进的方法一种是使用动态逻辑.动态逻辑引入了动态连结词,如[],⟨⟩等.这样,程序S的部分正确性问题就可以表示为$P\rightarrow[S]Q$,即P真蕴含当S终止时Q为真.如果S是确定性语句,则其完全正确性问题可以表示为$P\rightarrow\langle S\rangle Q$,即$P$真蕴含$S$在某一时刻终止且$Q$为真.与上述方法对应的另一种方法是把时态逻辑引入到程序逻辑中,以刻画程序的全部行为,即把程序的执行过程看成机器状态的一个变化序列.

程序逻辑的研究为公理语义学和程序验证提供了理论基础,另外也为软件工程中的功能描述和自动化程序设计提供了有力工具.

霍尔逻辑(Hall logic) 即“程序逻辑”.

形式语义学(formal semantics) 程序设计理论的组成部分.它以数学为工具,利用符号和公式,严格地定义程序设计语言的语义,使语义形式化.程序设计语言用于书写计算机程序,和自然语言一样,有其语法、语义和语用等三个方面的因素.程序设计语言的语法是指程序的组成规则,表示程序的结构或形式;其语义是指程序的含义;其语用是指程序与使用者的关系,表示程序的使用效果.通常的程序设计语言的语法是规定程序组成方法的一些规则,称为具体语法;而在形式语义学中使用一种讨论程序分解的语法规则,称为抽象语法.抽象语法采用的记号和表示方式是各种不同的程序设计语言的具体语法的一种抽象,它适用于一切程序设计语言.形式语义学中使用元语言定义程序设计语言的语法或语义.计算机程序规定对计算机系统中数据的加工过程.形式语义学的基本方法是把程序加工数据的过程及其结果形式化,从而定义程序的语义.按形式化中侧重面和使用的数学工具的不同,形式语义学可分为以下四个分支:

1. 操作语义学.着重模拟数据加工过程中计算机系统的操作.

2. 指称语义学.主要刻画数据加工的结果,而不是加工过程的细节.

3. 公理语义学. 用公理化的方法描述程序对数据的加工.

4. 代数语义学. 把程序设计语言看做是刻画数据和加工数据的一种抽象数据类型,使用研究抽象数据类型的代数方法,来描述程序设计语言的语义.

操作语义学(operational semantics) 形式语义学的一个分支. 大体上是以直接叙述“如何做”来定义语言语义的,即将程序设计语言中各个成分所对应的计算机系统的操作来作为语言成分的语义. 操作语义学的基本思想来源于程序设计语言的实施. 程序设计语言的实施是在具体的计算机系统中按照语言的语义编制语言的翻译程序,将语言中各个成分翻译成计算机系统中相应的一组操作. 由于语言的语义应该是标准的,不应依附于一个特定的计算机和一种具体的实施,故操作语义学中引进了抽象机器的概念,用抽象机器的指令来解释语义.

1964 年 1 月,英国人兰丁(Landin,P. J.)使用“栈-环境-控制-外储”抽象机器(简称 SECD 机器),第一次系统而严格地陈述了表达式的操作语义. IBM 公司的维也纳实验室提出描述操作语义的一种元语言——维也纳定义语言(简称 VDL). 1974 年,欧洲计算机制造商联合会和美国国家标准局正式建议使用 VDL 定义的 PL/1 语义作为 PL/1 的标准. 1980 年,英国爱丁堡大学提出结构式操作语义学,在一般的数学结构(不必是抽象机器)上用数学的归纳关系建立语义的解释系统. 在定义并发式程序设计语言的语义时,其他语义学方法都遇到了困难,而结构式操作语义学却显示了它独有的长处,对语言的设计和实施有其指导作用.

指称语义学(denotational semantics) 形式语义学的一个分支. 是以程序设计语言成分的指称物作为语言成分的语义的. 指称语义学方法认为,程序设计语言中各个成分的语义应该是执行语言成分所要得到的最终效果,而效果不因计算机系统的不同而改变. 执行语言成分产生的最终效果被看做是语言成分的所指,称为语言成分的指称物.

指称语义学方法把语言中的每一成分与一个数学对象相对应,称为从前者到后者的映像,后者即称为前者的标志(即指称),因此,指称语义学亦称标志语义学. 指称语义学方法在定义语言的语义时,先确定指称物,然后给出语言成分至指称物的语义映像,这个映像要求满足:每个成分都对应有指称;复合成分的指称只依赖于它的子成分的指称. 指称语义规定,这种映像是层次结构的,即若语言成分 a 由一些子成分 $a_1,a_2,\cdots,a_n$ 复合而成,则 a 的指称应完全用 $a_1,a_2,\cdots,a_n$ 的指称表示出来. 映像函数往往是递归的,这种递归函数的定义,可用最小不动点理论来确定. 指称语义已成功地用来定义了 CHILL,ADA 等大型语言以及数据库、操作系统等. IBM 公司的维也纳实验室设计的元语言 META Ⅳ 提供了用指称语义描述软件的工具. 程序设计语言的指称语义独立于语言的实施,便于语言设计者用来规定语言的语义. 指称语义学在未来的程序设计语言中将发挥更大的作用.

公理语义学(axiomatic semantics) 形式语义学的一个分支. 不同的人在了解程序的含义时有不同的要求. 公理语义学方法就是研究如何将这些不同的要求形式化,并根据这些要求严格给出程序设计语言的有关语义. 该方法是在一阶谓词演算的基础上进行如下扩充:

1. 一组公理. 其形式为 $P\{S\}Q$,这里 P 和 Q 是谓词公式,S 是一组动作(例如,某个程序设计语言中的一组语句). 其含意为如果在执行 S 前 P 成立,则执行 S 后 Q 成立.

2. 一组推导规则. 其形式为

$$\frac{P_1\{S_1\}Q_1, P_2\{S_2\}Q_2, \cdots, P_n\{S_n\}Q_n}{P\{S\}Q},$$

其含义为:如果横线以上的都是公理或定理,则横线以下的也是定理.

这种公理系统已用来描述 PASCAL 和 EUCLID 等语言的语义. 但它是不完备的. 公理语义学是程序正确性研究的理论基础,程序验证的研究也进一步促进公理语义学的发展.

代数语义学(algebraic semantics) 形式语义学的一个分支. 用代数方法研究计算机语言的语义. 该方法以数据类型抽象化为其主要特征,以非齐性代数及代数间的同构映射为其主要工具. 它把计算机语言形式地定义为满足某种公理体系的抽象代数结构,然后利用这种代数结构的性质来证明用该语言编写的程序的正确性. 当把一个程序设计语言看做是抽象数据类型时,就可用代数方法来描述它的语义. 具体做法是:使每个语法符号对应 S(表示由一组称为类子的元素构成的集合)中的一个类子,每个语法规则对应 O(表示由一组运算符构成的集合)中的一个运算. 又使语义关系对应公理系统 E. 为了表达上下文条件,可把同态映射扩充为弱同态,即允许同态映射由部分函数实现. 在一定的条件下,弱同态意义下的终结代数是存在的,并等价于程序的最小不动点语义.

程序正确性理论(theory of program correctness) 程序设计理论的一个重要组成部分. 研究如何使用程序设计语言编制程序,以正确实现预定的目标. 程序正确性理论提出编制正确程序的两种途径. 一种称为程序验证,研究如何使用数学推理来严格论证程序是否符合其目标. 另一种称为程序综合,研究如何由给定目标出发,逐步构造一个在计算机

系统上可运行的程序，而且要求构造过程的每一步都是严格保持正确无误的. 在保证正确性的前提下，为提高程序运算效率而实现的程序之间的转换，称为程序变形，也是程序正确性理论的一个重要内容. 程序综合和程序变形是自动程序设计的理论基础.

程序验证(program verification)　研究程序正确性的理论. 即研究如何使用数学方法严格证明一个程序是符合其预定的目标的，因而是正确无误的. 程序验证与程序测试有别，程序测试只能发现程序的错误，而不能确保程序没有错误.

早在 1947 年，冯・诺伊曼(von Neumann, J.)就发表论文提到程序正确性证明. 1967 年，美国的弗洛伊德(Freud, G.)系统地提出验证程序正确性的归纳断言法. 1969 年，英国的霍尔(Hall, C. A. R.)将归纳断言法形式化，提出程序验证的公理系统. 20 世纪 70 年代以来，又出现了能辅助用户正确编制程序的实用的半自动化程序验证系统，其中波伊尔-莫尔程序验证系统最为著名. 要验证程序，首先把程序所要实现的目标形式化，即使用数学公式表达程序加工的初始数据的范围(称为“输入谓词”)和程序加工的结果(称为“输出谓词”). 使用归纳断言法，只能证明在输入谓词成立的前提下，程序终止时输出谓词也成立. 但不能证明在输入谓词成立时，程序一定能终止. 不讨论程序终止性的程序验证，称为程序部分正确性的验证；包括终止性的验证，称为程序完全正确性的验证. 程序验证技术除了用于证明程序的正确性外，还可以从程序正确性角度评价程序设计方法和程序设计语言的优劣.

计算机体系结构

数字计算机(digital computer)　一种处理数字编码信息的计算机. 数字编码信息是用不连续的形式表示的数据，通常由“0”或“1”组成的代码来表示. 数字计算机中控制计算机工作的控制信息，也由数字形式表示. 它通常由一个或多个相互关联的运算处理单元、存储单元以及外部设备所构成，由存储在内部的程序加以控制. 由于已大量用于各种科学研究与计算、生产过程控制和管理及国民经济的各有关部门的管理、统计、决策等方面，甚至已进入人们的日常生活之中，故数字计算机当前已简称计算机.

模拟计算机(analog computer)　一种对模拟量进行运算和处理的计算机. 模拟量是一种连续变化的物理量，例如电流、电压、电阻、转动角速度、长度等. 在模拟计算机中处理的一切物理量，都由连续变化的电流或电压来表示. 模拟计算机的主要运算部件是由数运算放大器为基础构成的各种线性和非线性运算部件(如加法器、反相器、微分器、积分器、比例器、函数发生器等). 工作时根据被仿真的物理系统的数学模型用排题板将多种运算部件连结起来，在控制部件控制下可获得解答. 与数字计算机相比，模拟计算机具有响应速度快、接口简单、使用方便等优点，但计算精度、工作稳定性和可靠性，特别是应用的广泛性都难与数字计算机相比.

混合计算机(hybrid computer)　一种计算机. 是利用模拟量和数字量两种表示方式进行数据处理的电子计算机. 兼有模拟计算机和数字计算机二者的特点. 常用于统计分析、解偏微分方程、解最优化问题以及各种仿真系统中. 混合计算机通常有以下两类：混合模拟计算机和组合式混合计算机. 混合模拟计算机是一种以模拟技术为主，附加一些数字设备的混合计算机. 它与普通模拟计算机的主要区别是：

1. 运算部件频带加宽了.
2. 增加了模拟存储和逻辑部件.
3. 增加了可灵活连结的数字部件.
4. 控制部分增加了时序部件.
5. 增设了自动排题运算系统.

混合计算机利用数字逻辑部件实现各种控制信号间的布尔代数关系，因此能按预定的逻辑去控制模拟开关、存储部件及运算部件，以实现自动采样、自动设置和改变参数、自动改变程序、自动判决、自动迭代运算等. 组合式混合计算机是由模拟计算机和数字计算机加上相应的接口装置组成的一种混合计算机. 其中数字计算机的作用是：

1. 管理模拟计算机. 按预定程序控制模拟计算机工作，存储及处理模拟计算结果，编制各种输出记录.
2. 提高系统效率. 即编制各种混合软件，实现模拟部分的自动程序设计、设置及自动操作组合式混合计算机等.
3. 分担混合计算机中的部分计算工作. 即产生多变量函数，进行精确计算等.

ENIAC 计算机(electronic numerical integrator and calculator computer)　世界上第一台电子计算机. 全称为电子数字积分计算机. 1946 年 2 月由美国宾夕法尼亚大学莫尔学院制成，设计者为莫奇利(Mauchly, J. W.)和埃克特(Eckert, J. P.). 这台计算机使用了 18800 只电子管，占地面积 150 平方米，重达 30 吨. 加法速度为 5000 次/秒，乘法速度为 56 次/秒，比继电器计算机约快 1000 倍，比人快 20 万倍. 机器字长为 10 位(采用十进制)，内存仅有 20 个寄存器，不能存放程序，靠接插板编程序. 它在美国陆军弹道研究实验室运行了近十年.

计算机系统(computer system)　由计算机硬件系统和软件系统组成的机器系统. 硬件是系统工

作的物理实体，软件控制指挥硬件工作，扩展系统功能. 硬件由中央处理器、主存储器、辅助存储器、输入-输出设备、电源等组成. 软件由系统软件和应用软件组成.

计算机系统根据其数据表示分为数字计算机与模拟计算机. 数字计算机处理的数据是离散的数字量，而模拟计算机的数据是连续变化的模拟量. 当前数字计算机已被简称计算机. 计算机系统根据其功能可分为通用机与专用机. 前者兼顾科学计算、事务处理和过程控制等应用，后者按特定用途设计. 计算机系统按其性能和规模又可分为巨型计算机、大型计算机、小型计算机和微型计算机. 计算机系统按工作方式和系统结构分为：流水线处理机与并行处理机；多处理机与多机系统、分布式系统和计算机网；冯·诺伊曼机与非冯·诺伊曼机.

巨型计算机(supercomputer)　一种超级计算机. 通常指在一定时期内速度最快、性能极高、技术最复杂、规模最大、耗资极多的计算机. 它是一个相对的概念，一个时期内的巨型计算机到下一时期可能成为一般的计算机. 现代巨型计算机按照体系结构的发展可分为四代巨型计算机，即 SIMD 阵列机、向量机、并行多处理机和大规模并行处理系统. 现代的巨型计算机主要用于核物理研究、核武器设计、航空航天飞行器设计、国民经济的预测和决策、能源开发、气象预报、卫星图象处理、情报分析等各种科学研究，是强有力的模拟和计算工具.

大型计算机(large scale computer)　一种通用计算机. 是性能超过小型计算机而低于巨型计算机的一类计算机. 大型计算机为了获得高性能，一般都采用如下一些新技术：

1. 普遍采用大规模和超大规模集成电路，大大缩小了体积.

2. 为了解决逻辑门之间的延迟，提高整机系统性能，而采用高密度芯片包装新技术.

3. 采用空气冷却新技术以提高系统工作效率.

4. 采用向量处理机以扩大大型计算机的应用领域.

5. 采用新的内存工作原理以提高系统性能.

6. 采用流水线和交叉存取技术.

7. 采用容错与冗余技术.

大型计算机都配有极为丰富的各种系统软件和应用软件.

小型计算机(minicomputer)　一类计算机. 它是 20 世纪 60 年代中期发展起来的，规模介于大型计算机和微型计算机之间的一类计算机. 相对于大型计算机而言，小型计算机的软件、硬件系统规模比较小，但价格低、可靠性高、便于使用和维护. 初期的小型计算机指令功能比较简单，只配有汇编语言等初级软件. 经过 30 多年的发展，上述一些指标均有大幅度提高，其指令功能、存储容量、输入输出能力等方面都能与大型计算机系列中的低档机相匹敌，且性能价格比要高得多.

超级小型计算机(super-minicomputer)　一类计算机. 是 20 世纪 70 年代后期从典型的 16 位小型计算机基础上发展起来的不小于 32 位的一类小型计算机. 由于其性能大大超过了传统的小型计算机，同时又保持了小型计算机的价格和特点，所以被称为超级小型计算机. 它具有以下特点：

1. 具有 32 位系统结构.

2. 硬件直接支持高级数据类型.

3. 具有相当高的处理速度.

4. 具有虚拟存储器.

5. 具有总线结构.

6. 具有联网能力.

7. 具有通用多功能交互式操作系统.

8. 具有丰富的软件产品.

微型计算机(microcomputer)　一类计算机. 是用大规模集成电路或超大规模集成电路制成的微处理器、存储器和配套的输入输出接口等组成的计算机. 按不同的应用要求，配置相应的外围设备和软件，可以形成完整的微型计算机系统. 微型计算机是微电子技术发展到大规模集成阶段的产物，它的特点是体积小、重量轻、耗电少、可靠性高. 随着集成电路集成度的提高而价格降低，微型计算机的性能价格比按指数上升，应用日益广泛，已成为工厂自动化、办公自动化和家庭自动化的基础.

微型计算机是体积更小的计算机，按其结构不同，可分为单片机、单板机和多板机三类. 按其用途不同，可分为通用与专用两类. 通用微型计算机的指令系统，一般可用于数值计算和信息处理等许多方面，配有比较完善的系统软件和外围设备. 专用微型计算机的逻辑结构根据具体算法特点而进行设计，用以满足快速响应要求. 这两类机器用于信息作业的物理形式转换、处理和传输等各个环节. 微型计算机按其体积大小又有台式、便携式和掌上型等几类. 微型计算机可以独立工作，也可按终端仿真方式作工作站使用. 随着微型计算机的处理能力及存储容量的提高和小型计算机尺寸的缩小，微型计算机和小型计算机之间的差异正在逐步缩小.

工作站(workstation)　计算机的一种类型. 工作站是 20 世纪 80 年代出现的一种高性能的、可放在办公桌上的 32 位微型计算机. 它比个人计算机有更强的能力和更高级的图形功能、更灵活的联网能力、更佳的性能价格比. 当前工作站处于个人计算机和更大的小型计算机或大型计算机之间，但它们之间的界限正变得日益模糊. 工作站常用于处理个人

计算机所不能实现的复杂任务，否则需要使用价格贵得多的大型计算机或小型计算机系统．工作站采用超级小型计算机和大型计算机的系统结构．

1．具有虚拟存储系统．能支持多任务的分时运行，但其强有力的单用户多任务环境能在任何时候向每个用户提供得到保证的处理能力．

2．可采用多处理器结构，实现并行运算．

3．配有网络接口，并有支持通信和分享资源的软件．能灵活地与其他工作站上的用户连成局部网络，实现有效通信和资源共享．

4．工作站一般设有图形控制部件，用以解释并执行级别较低的图形命令．

多数工作站还配置有图形处理机，以解释和执行级别较高的图形命令，实现坐标变换、浓淡、隐藏、剪裁、填充和比例缩放等功能，大大提高了图形处理速度．图形软件不仅使工作站能提高一般的图形显示，还提供分辨率很高的彩色交互式图形显示．像多窗口显示、菜单提示选择、图形编辑以及图标等新的功能和概念也已经被引入了人机接口．目前，工作站大量用于 CAD/CAM 领域以及软件工程和人工智能等领域．

多媒体计算机（multimedia computer） 能处理多种媒体信息的计算机．具有对声音、文字、图形、图象、动画等多种媒体信息进行获取、编辑、存储、检索、展现、传输等处理能力的计算机称为多媒体计算机．当前多媒体计算机多指多媒体工作站和多媒体 PC 机，多媒体 PC 机缩写为 MPC．

1991 年 11 月，美国微软公司与其他多媒体产品制造商一起制订了一个关于多媒体 PC 机 Lever Ⅰ 的规格．其硬件指标为：

80386 SX 以上的处理器，主频≥16 MHz；

2 MB 以上内存，30 MB 以上硬盘；

VGA 显示（16 色，最好 256 色）；

声卡（8 位以上 AD/DA 量化精度，11.025 kHz 以上采样频率，带有 MIDI 音乐合成器，占用 CPU 开销＜20％）；

CD-ROM 驱动器（平均访问时间≤1s，数据传输速率≥150 kB/s，占用 CPU 开销＜46％，可播放数字音乐）；

MPC Lever Ⅰ 不要求配置专用的视频处理硬件，需要处理视频信息时，使用软件完成．

1993 年 5 月，美国多媒体市场委员会又发布了 MPC Lever Ⅱ 的性能规格，其硬件指标比 Lever Ⅰ 明显提高：

80486 SX 以上的处理器，主频≥25 MHz；

4 MB 以上内存，160 MB 以上硬盘；

VGA 显示（640×480，65536 色）；

声卡（16 位的量化精度，44.1 kHz 采样频率，带 8 音符复音的 MIDI 合成器，支持 CD-ROM×A 声频功能，支持 IMA 推荐的 ADPCM 算法）；

CD-ROM 驱动器（平均访问时间为 0.4s，数据传输速率达 300 kB/s，符合 CD-ROM×A 规范）；

视频处理功能（在 40％CPU 带宽的情况下每秒可传输 1.2 兆像素）．

多媒体计算机软件包括多媒体设备的 I/O 控制软件、操作系统的多媒体功能扩充软件和多媒体编辑工具与写作工具软件等．家用 PC 机通常都具有多媒体功能．

冯·诺伊曼计算机（von Neumann machine） 根据存储程序原理构造的计算机．存储程序原理由冯·诺伊曼（von Neumann, J.）于 1946 年提出．其要点是：

1．数据和指令都由二进制编码表示．

2．计算机由输入设备、输出设备、存储器、运算器和控制器组成．

3．存储器按地址进行访问，地址结构是线性编址的一维结构．

4．数据和为实现数据加工处理而编制的程序一起存入存储器．

5．采取集中的统一控制．

计算机的工作过程就是依次、逐条从存储器取出程序中的指令并执行指令的过程．指令的执行一般是顺序的，但可由转移指令实现程序的分支．指令是计算机硬件能直接识别和执行的操作命令，指令一般由操作码和地址码组成．操作码规定指令所执行的操作，而地址码则提供操作数的地址．将要取出执行的指令在内存中的地址由控制器中程序计数器提供．综上所述，传统的冯·诺伊曼计算机实质上是一种由指令流驱动的计算机，目前大多数计算机仍属于冯·诺伊曼计算机．

计算机组织（computer organization） 计算机系统结构的逻辑实现和物理实现．研究和建立计算机运算器、存储器、控制器、输入设备和输出设备等基本功能部件的组成、相互连结和相互作用，实现计算机在机器指令级的功能和特性，是逻辑实现的内容．进而用各种元件、器件、逻辑部件实现其逻辑功能则属于物理实现的内容．尽管现今大多数计算机组织，就其结构原理来说，仍是以存储程序原理为基础的冯·诺伊曼计算机，但在计算机组织的各方面都出现了许多新的特点．运算器组织已由单一化向多功能化发展．除原单一的 ALU 外，还可包括乘法器、除法器、浮点运算器，甚至将数值协处理器包括在内．控制器由单指令串行执行向多指令的重叠执行、流水线结构发展．指令系统则有复杂指令集（CISC）与精简指令集（RISC）的融合．存储器的组织则向层次结构发展．在中央处理器内或与主存储器

之间均设置高速缓冲存储器，在主存储器和辅助存储器一级组成虚拟存储器．输入-输出接口则向标准化发展．在单机的基础上进一步向多机并行处理系统发展，出现了向量处理机、阵列处理机、分布式处理系统、数据流处理机等．

计算机数据表示（computer data representation） 计算机对处理对象的描述方法．是便于计算机硬件识别、运算、处理、存储、传送的数据表示方法．它涉及数的进位计数制、定点或浮点表示、正负数的符号表示，以及数的原码、反码、补码表示及其相应的运算方法．

计算机中常采用二进制，这是因为：

1．二进制中每一位数的符号只有 0 和 1 两个，易于由具有两个稳定状态的物理器件来实现．例如，用晶体管的饱和导通状态表示 0，则晶体管的断开状态可表示 1．

2．二进制数的运算方法简单．

3．二进制数的算术运算易于转化为逻辑操作来实现，因此运算器可用逻辑电路实现，且便于使用逻辑代数简化处理机的逻辑设计（参见“基数计数法”）．

计算机中的数据符号通常用 0 表示正，1 表示负．例如，十进制数＋6 可表示为二进制数的 0110，－6 可表示为 1110，其中最高位表示符号，称为符号位．

计算机中数据按小数点位置固定与否分为定点数和浮点数两种．定点数的小数点通常固定在符号位后的称为定点小数，小数点固定在最低数值位后的称为定点整数．小数点在计算机中是隐含表示的．定点数的表示范围窄，使用不便，但易于硬件实现．浮点数的小数点位置不固定，而是由阶码指定．浮点数的表示范围比定点数宽、精度高，但算法复杂，硬件实现不易（参见“定点数的表示法”和“浮点数的表示法 ”）．为了简化加减运算方法，计算机中的数又有原码、反码和补码表示法，其中最常使用的是补码（参见“原码”、“反码”、“补码”等）．

基数计数法（radix notation） 亦称按位计数法或进位计数法．计算机的计数方法之一．是以基数和权来表示的计数方法．基数用来表示符号的个数；权是与该数位所在位置有关的因子，一般为基数的整数次幂．计算机中常用的基数计数法有以下四种：

1．二进制计数法．基数为 2，所用的数字符号为 0，1，权是 2 的整数次幂，进位方式是“逢二进一”．

2．八进制计数法．基数为 8，所用的数字符号为 0，1，2，3，4，5，6，7，权是 8 的整数次幂，进位方式是“逢八进一”．

3．十进制计数法．基数为 10，所用的数字符号为 0，1，2，3，4，5，6，7，8，9，权是 10 的整数次幂，进位方式是“逢十进一”．

4．十六进制计数法．基数为 16，所用的数字符号为 0，1，2，3，4，5，6，7，8，9，A，B，C，D，E，F，其中 A，B，C，D，E，F 分别表示十进制数的 10，11，12，13，14，15．其权是 16 的整数次幂，进位方式是“逢十六进一”．

按位计数法（radix notation） 即“基数计数法”．

二-十进制计数法（binary-coded-decimal） 亦称二-十进制（BCD）．计算机的计数方法之一．是二进制编码的十进制．用四个二进制位的数值直接表示一个十进制数据位的编码，称 BCD 码．BCD 码又分为 8421 码、2421 码余 3 码、格雷码等多种，但最常用的是 8421 码，8，4，2，1 分别表示四位二进制数中自高到低各位的权值．这种码制从 0 至 9 的表示法如表所示．

十进制	0	1	2	3	4	5	6	7	8	9
8421 码	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001

从表中可知，这种方法对于只有 1 位的十进制数，其结果和二进制相同，但如果是两位以上的十进制数就不同了．例如，十进制数 12，用这种码制写出时有 8 位，即 00010010，而不是二进制的 1100．这种计数法既直观又简单，常被采用．

定点数的表示法（fixed-point representation） 一种计算机数据表示方法．数的表示采用固定小数点位置的形式，即在处理机的运算器中，数据小数点的位置是隐含固定的．通常，小数点隐含固定在数据最右端的，称定点整数，如 101；小数点隐含固定在数据最左端的，称定点小数，如 0．1001．定点数的表示法会给用户带来很多麻烦和限制：

1．用户的初始数据、中间结果或最后结果可能在很大的范围里变化，程序员不得不在运算的各个阶段预先引入比例因子，把数据统一放大或缩小．

2．一定长度的定点数据所能表示的数据范围和精度是很有限的．

浮点数的表示法（floating-point representation） 计算机数据的一种表示方法．数据的小数点位置不固定，而是浮动的．浮点数表示法中，数 N 可表示为

$$N = \pm d \times 2^{\pm p},$$

式中 d 称为 N 的尾数（或数值），p 称为数 N 的阶码（或指数）．d 是一个纯小数，p 是一个整数．p 的符号表示小数点浮动的方向．p 为正，表示小数点右移 p 位；p 为负，表示小数点左移 p 位．例如

$$N = +0.1101011 \times 2^{+6},$$

其中阶码为＋6，d 的小数点应右移 6 位，即

$$N = +0110101.1.$$

又如

$$N = 0.1101011 \times 2^{-6},$$

则小数点应向左移动 6 位，即

$$N = 0.0000001101011.$$

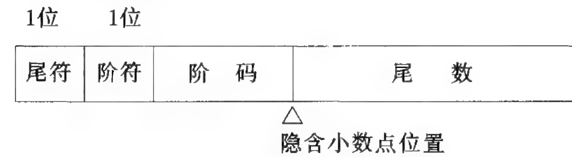
在浮点数的表示法中，同一数 N 有不同的表示形式，其中尾数的最高位为有效数值 1 的浮点数形式称为规格化浮点数. 例如 $N=110.1101$ 表示成尾数 8 位的浮点数形式时有

$$\begin{aligned} N &= 0.1101101 \times 2^{+3} \\ &= 0.0110110 \times 2^{+4} \\ &= 0.0011011 \times 2^{+5} \\ &= 0.0001101 \times 2^{+6} \end{aligned}$$

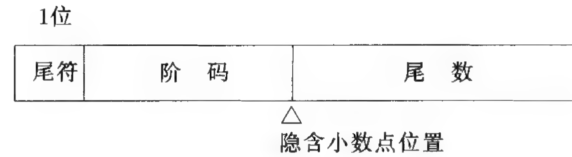
等形式，但 $N=0.1101101\times 2^{+3}$ 仅是规格化形式，有效数值位最多，精度最高.

采用浮点数表示法的优点是可以用位数较少的信息代码来表示绝对值很大的数，这样用同样位数的信息代码所表示的数，其浮点数表示法的表示范围比定点数表示法的表示范围大很多，使用更方便，但其运算方法比定点数表示法复杂.

在计算机中浮点数的表示格式为：



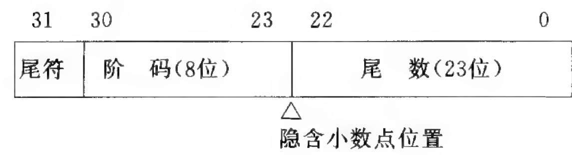
或



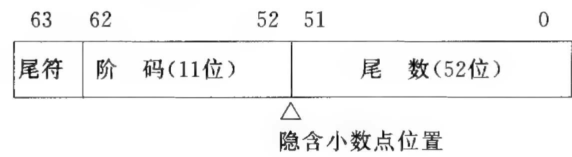
其中阶码用移码表示时，阶符隐含在阶码内.

目前计算机中，常用的浮点数表示格式有三种，即根据 IEEE 标准推荐的 32 位格式和 64 位格式以及 Intel 公司使用的 80 位格式：

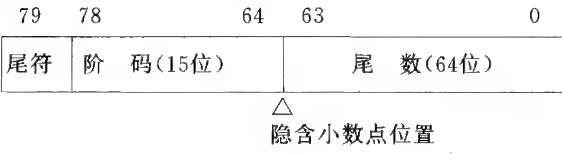
1. 32 位格式：



2. 64 位格式：



3. 80 位格式：



原码(true form) 电子计算机中数的一种表示形式. 数的原码形式除掉其符号表示位外与原来数的绝对值形式一样，只是数的正负符号分别用“0”和“1”表示. 例如，＋46 的二进制数原码表示为 00100110，而 －46 的二进制数原码表示为 10100110(第一位为符号表示位). 数值“0”则有两种原码表示：“＋0”为 00000000；“－0”为 10000000. 原码表示法对于数的乘除运算相当方便，但对加减运算则较复杂. 在两个数进行加减运算时，需要判定它们的符号和绝对值大小才能确定操作关系和操作类型. 为了简化处理机的加减运算，通常很少采用原码表示法，而是采用二进制补码或二进制反码表示数据.

反码(radix-minus-one complement) 亦称基数减 1 补码. 计算机中数的表示形式之一，是基数计数制中以基数减 1 后分别减去每一位数字后所得到的数. 例如，二进制计数制中，基数是 2，基数减 1 为 1，则 －0101101 的反码是 1111111 － 0101101 ＝ 1010010；十进制计数制中，基数是 10，基数减 1 为 9，则 －0543 的反码是 9999－0543＝9456. 对于正数，二进制反码的表示式和原码相同；对于负数，其二进制反码表示式是将二进制数值表示式中的 0 改为 1，1 改为 0，再冠以负号 1. 例如 00101110 的反码仍为 00101110；10101110 的反码为 11010001.

基数减 1 补码(radix minus one complement)即“反码”.

补码(complement) 计算机中一种数的表示形式. 在基数计数制中，从基数减 1 后分别减去某数的每一位数字，然后再加 1 所得到的数. 对于正数，其二进制补码的表示式和原码相同；对于负数，其补码表示式是将其反码的最低位再加 1 而得到. 例如，二进制数 00101110 的补码仍为 00101110；10101110 的反码是 11010001，其补码为 11010001 ＋1＝11010010. 在补码表示式中，数值 0 的补码是惟一的，即 00000000(全零). 补码表示法对正负数的加减运算相当方便. 利用补码可使减法运算转化为加法运算，这样在计算机中就可用加法器来进行加减运算.

位(bit) 量度信息的单位. 二进制数据或代码的每一数位(0 或 1)称为“位”，是计算机信息的最小单位. bit 是英文 binary digit 的缩写，故位的音译名为“比特”. 二进制数的一位包含的信息量称为一

比特.

字节(byte) 度量数据长度的一种常用单位.通常用八位二进制数组成一个字节,一个键盘上的字母、数字或其他符号用一个 ASCII 码表示,一个字节可容纳一个 ASCII 码(含一位奇偶校验位).计算机存储器的常用编址单位是以字节为单位的.计算机的通信传输单位一般也以字节为基础.

字(word) 计算机数据的表示单位之一.指一组字符或一组二进制位.在计算机系统中,在存储、传送或操作时,字是作为一个整体单元的一组字符或一组二进制位.字的重要属性是字长,即组成一个字的二进制的位数或字符的数量.字长有固定字长和可变字长两种.字通常由若干字节(参见"字节")组成,因此字长常为 8 的倍数,即 8 位、16 位、32 位、64 位等,以便于处理字符和十进制数据.存储器中,地址单元可按字节或字编址.每个字都是可寻址的.字还涉及下述一些问题:

1. 数据的数值范围.
2. 数据的精度.
3. 指令的直接寻址能力.
4. 指令格式.
5. 硬件规模.
6. 容错能力.
7. 硬件成本等.

ASCII 码(American standard code for information interchange) 美国信息交换标准代码.是美国标准协会向国际标准化组织建议的美国标准代码.由 8 位代码组成,其中 7 位为数据位,最高位为奇偶校验位.用 7 位二进制数编码,可表示 128 个不同的字符,包括大写和小写的字母、数字、标点、各类符号和各种控制字符.数字代码和字母代码是有序的,可以利用代码的数值比较对字符序列进行字典排序.ASCII 码是目前最流行的一种字符数字代码,已广泛应用于数据处理、通信系统等设备的信息交换.

计算机硬件(computer hardware) 计算机系统中的一切物理装置.包括各种电子、机械、光电器件、磁性器件等元件及其组成的各种部件及设备.主要部件和设备包括运算器、控制器、主存储器、辅助存储器、输入-输出设备、总线、电源等.这些部件和设备按照计算机组织结构原则构成的有机整体称为计算机硬件系统,为计算机软件的运行提供了必要的支持.

处理机(processor) 计算机的一个核心部件.由中央处理器、主存储器、输入-输出接口、总线组成的能执行程序、按程序的要求执行指令序列并完成处理任务的部件.处理机处理能力可用以下的几种指标或参数来度量:

1. MIPS——每秒执行的百万条指令数.
2. MFLOPS——每秒能给出百万个浮点处理结果数(主要用于向量处理机).
3. PDR——执行每条指令传递的平均位数与指令处理平均速率的乘积.

早期处理机的结构是以运算器为中心,而当今都是以主存储器为中心.处理机与外围设备的数据传递控制方式有:

1. 程序直接传递控制方式.
2. 程序查询方式.
3. 程序中断传递控制方式.
4. 直接存储器存取方式(DMA 方式).
5. 通道方式.
6. I/O 处理机方式.

随着计算机系统结构的发展和超大规模集成电路技术的进步,已能构造不同结构、不同功能、不同用途的多种处理机,如阵列处理机、向量处理机、数组处理机、输入输出处理机(I/O 处理机)、数据库处理机等.

芯片(chip) 计算机及所有智能硬件的一种基本组成器件.它通常是在硅晶片上,采用半导体超大规模集成电路工艺,把成千上万、甚至数百万个晶体管元件与连线集成在微小的芯片上,获得强大的电路功能,从而使计算机及各种智能装置实现微型化.芯片主要有微处理器芯片、存储器芯片、系统控制芯片集、超级 I/O 芯片以及特殊用途芯片等.1970 年,英特尔(Intel)公司生产了第一个 1kB 容量的存储器芯片.1971 年,推出第一个 4 位微处理器芯片 4004.接着推出 8 位芯片 8008.1981 年的准 16 位芯片 8086 成为 PC 机的心脏.1982 年的 80286,以及后来的 80386,80486,为 PC 机性能的发展做出了贡献.1993 年,英特尔推出新的奔腾(pentium)芯片,使微处理器技术进入新阶段.奔腾芯片是最重要的一类微处理器芯片.1993 年的经典奔腾集成 310 万个晶体管,运行速度是 4004 的 1500 倍,达到 90MIPS.不久又有多能奔腾、高能奔腾问世,它们通称为奔腾 I 代.1997 年推出奔腾 II 代,集成 750 万个晶体管.1999 年 2 月推出奔腾 III 代,集成 950 万个晶体管,时钟频率为 450～1000MHz.2000 年 11 月推出奔腾 IV 代,集成 1200 万个晶体管,时钟频率为 1.4～1.5GHz,使处理音频、视频、上网及 3D 图形显示的速度更快.奔腾采用的计算机技术有:超标量技术、超流水线技术、把指令 cache 与数据 cache 分开的哈佛结构、分支动态预测、固化常用指令、拓宽数据总线及突发模式、采用 PCI 局部总线、错误检测与校验技术、电源效率技术、支持多重处理等.

指令(instruction) 计算机科学术语.指规定

计算机操作类型及操作数或操作数地址的一组字符.不同计算机都有各自的指令.指令都用二进制形式表示,常称这些指令代码为指令码.它仅能被各自的中央处理器理解和执行.指令码分为操作码和地址码两部分.操作码规定要执行的操作类型,如加、减、传送、条件转移等.地址码部分指定操作数所在地址(源地址)和操作结果要送往的地址(目的地址).按照寻址方式的不同,它们可以是操作数、操作数地址或其他获得操作数所需要的其他地址信息.指令按地址码域中所包含的地址数目而分为三地址指令、二地址指令、单地址指令和零地址指令.按指令长度可分为单字节指令、双字节指令和多字节指令等.按指令操作类型可将指令分为算术逻辑指令、传送指令、程序转移和控制指令、输入和输出指令、特别指令等.

指令系统(instruction set) 一台计算机的CPU能直接识别和执行的全部指令的集合.指令系统就是机器语言,一条指令就是一个语句.机器指令语句所满足的语法规则就是指令的结构格式.指令系统是计算机性能广泛而集中的表现,是计算机设计的重要方面.从系统设计的角度看,指令系统是软件与硬件的界面.硬件设计人员要根据指令系统的约束条件,构造硬件组织,由硬件支持指令系统的功能的实现.软件设计人员在指令系统的基础上建立程序系统,扩充和发挥机器的功能.

精简指令集计算机技术(reduced instruction set computer technique) 简称RISC技术.一种计算机指令系统的设计技术.采用精简而有效的指令系统,从而使处理机内部结构简单、执行速度快的一种计算机指令系统设计技术.采用RISC技术的计算机称为精简指令系统计算机.RISC技术的主要设计思想是采用精简且能有效地支持高级语言的指令系统,尽量采用寄存器-寄存器操作,使指令格式力求一致.采用RISC技术可使处理机结构大为简化,再结合流水线技术和优化编译技术,就可以研制出结构简捷、集成度较低,但功能很强的处理机.RISC技术包括以下特点:

1. 机器语言指令数应尽可能少.
2. 指令寻址方式应尽量少.
3. 指令格式的种类应尽可能少.
4. 单周期操作,使一个机器周期能执行一条指令.
5. 只有存数和取数指令访问存储器,其余指令都只在寄存器间操作,所以CPU中通用寄存器相当多.
6. 硬布线控制逻辑为主,微程序控制为辅,以提高指令执行速度.
7. 优化编译,有效支持高级语言.

数字逻辑(digital logic) 数字电路的基本概念之一.是研究具有两个离散状态的数字电路及其组成的数字系统的逻辑功能、特性及其实现的理论和方法.对一个数字电路,研究其逻辑功能和特性称为逻辑分析.根据确定的逻辑功能和特性设计相应的逻辑电路称为逻辑综合.数字逻辑电路分为组合逻辑电路和时序逻辑电路两大类.输出状态仅取决于当前输入状态的电路称为组合逻辑电路;输出状态不仅取决于当前的输入状态,还与过去的输入状态有关的电路称为时序逻辑电路.时序逻辑电路又分为同步时序逻辑电路和异步时序逻辑电路.

中央处理器(central processing unit) 简称CPU.处理机中处理数据和执行指令的部件,由运算器和控制器组成.运算器在控制器控制下完成指令要求的数据处理.控制器控制程序的运行,依次提供指令在内存的地址,从主存读出指令、分析指令,产生执行指令功能所需的全部微操作控制信号,控制处理机内相应部件完成指令的功能.

一条运算型指令的执行过程包括从主存读出指令、分析指令、计算源操作数有效地址、从主存取操作数、计算目的操作数有效地址、从主存取目的操作数、按操作码要求执行数据运算、将运算结果存回目的操作数地址的过程.如果操作数在寄存器中,则寻址过程可以大大简化.

运算器的基本功能是算术运算、逻辑操作、移位等.常由算术逻辑运算部件ALU、移位器、输入数据逻辑器、锁存器、通用寄存器组等组成(参见“运算器”).控制器负责生成并提供指令地址,取出指令,分析指令,产生完成指令功能所需的一系列微操作控制信号,并送到相应部件完成指令的功能.在程序的执行过程中,如计算机系统发生故障,或有随机的输入、输出或其他意外事件发生时,中央处理器可采用中断方式,暂时中止现行程序的运行,输入执行中断处理程序,处理随机事件,处理结束后自动返回原来被中止的程序继续执行原程序.中断功能使中央处理器的功能提高.控制器由指令部件、地址形成部件、时序部件、微操作控制部件、中断控制逻辑和总线控制逻辑等组成.指令部件提供将要取出执行的指令地址,保存取出执行的指令,分析指令;地址形成部件计算操作数有效地址或转移地址,地址形成部件也可由ALU部件代替;时序部件提供执行过程中的同步定时信号;微操作控制部件根据指令分析的结果产生执行指令,完成指令功能所需的各种微操作控制信号,中断控制逻辑完成对随机异常事件的处理,总线控制逻辑用于对系统总线使用的管理(参见“控制器”).

随着计算机技术、超大规模集成电路技术的发展,中央处理器可集成在一个半导体芯片上,称为微

处理器，而且功能正在不断提高. 在新型的中央处理器芯片中可以包括多个整数运算部件、浮点运算部件、向量运算部件、指令流水部件、高速缓冲存储器、存储管理部件等.

运算器(arithmetic unit) 亦称算术逻辑运算部件，简称 ALU. 计算机中实现数据加工处理的部件. 运算器的基本操作包括加、减、乘、除、与、或、非、异或、移位、传送等算术运算和逻辑操作. 基本运算器由加法器、数据选择器、锁存器、暂存器、通用寄存器组、移位器以及连结各部件的数据通路及控制逻辑组成. 运算器执行各种操作所需的控制信号由控制器提供. 参与运算的数据由指令地址域指定，可来自存储器或寄存器，运算结果通常送回指定的寄存器或存储器. 根据操作数的进位计数制，可将运算器分为二进制运算器和十进制运算器；根据对操作数的各位采用串行逐位运算或同时运算，又可分为串行运算器和并行运算器；按照数据的表示格式，运算器又分为定点运算器和浮点运算器. 通常计算机中是二进制定点并行运算器. 为了提高运算处理速度和功能，微型和小型计算机中常在基本运算器外配置数值协处理器，随着超大规模集成电路技术的发展，现已将数值协处理器集成在中央处理器内，从而大大增强了中央处理器中运算器的性能. 某些大型计算机中常采用多个相同功能或不同功能的运算器，并引入流水线技术从而实现数据的并行处理功能，大大提高了运算速度.

算术逻辑运算部件(arithmetic logic unit) 即“运算器”.

控制器(control unit，controller) 计算机中央处理器的组成部件. 用以控制程序的运行. 控制器产生取指令、分析指令、取操作数和执行指令所需要的全部控制信号序列，送到运算器和其他功能部件以控制计算机各部件协调一致地工作，完成指令规定的功能. 控制器按其产生微操作控制信号序列的方法不同而分为硬连线逻辑控制和微程序控制两类. 硬连线逻辑控制主要采用组合逻辑电路或可编程逻辑阵列 PLA 产生微操作控制信号序列. 微程序控制则依靠执行存储在控制存储器中的与机器指令相对应的微指令序列(微程序)来产生执行该机器指令所需的全部微操作控制信号序列.

控制器的重要部件有指令寄存器、指令地址寄存器(常称为程序计数器或指令计数器)、指令译码器、时序部件、微操作信号产生器(微程序控制器中则用控制存储器、微指令地址形成部件、微指令地址寄存器、微指令寄存器及其译码电路代替)，以及中断和总线控制逻辑等组成. 中断控制逻辑使计算机能及时处理计算机运行过程中产生的各种异常事件，大大扩大了计算机系统的功能. 总线控制逻辑实现对总线使用权的合理调配和控制. 控制器的控制方式分为同步和异步两种. 同步控制方式是按照一定的节拍发出各种控制命令；异步控制方式是互锁应答方式，即前一操作结束发出的回答信号作为后继操作的启动信号. 同步控制方式常用于处理机内部控制，异步控制方式常用于 I/O 控制.

中断(interrupt) 计算机科学术语. 指处理机处理程序运行中出现的紧急事件的整个过程. 程序运行过程中，系统外部、系统内部或者现行程序本身若出现紧急事件，处理机立即中止现行程序的运行，自动转入相应的处理程序(中断服务程序)，待处理完后，再返回原来的程序运行，这整个过程称为程序中断；当处理机接受中断时，只需暂停一个或几个周期而不执行处理程序的中断，称为简单中断. 中断又可分为屏蔽中断和非屏蔽中断两类. 可由程序控制其屏蔽的中断称为屏蔽中断或可屏蔽中断. 屏蔽时，处理机将不接受中断. 反之，不能由程序控制其屏蔽，处理机一定要立即处理的中断称为非屏蔽中断或不可屏蔽中断. 非屏蔽中断主要用于断电、电源故障等必须立即处理的情况. 处理机响应中断时，不需执行查询程序. 由被响应中断源向 CPU 发向量地址的中断称为向量中断，反之为非向量中断. 向量中断可以提高中断响应速度.

程序中断(program interrupt) 见“中断”.

简单中断(simple interrupt) 见“中断”.

栈(stack) 亦称堆栈. 计算机内存中由程序指定的一个后进先出(LIFO)存储区或 CPU 内部的一个 LIFO 寄存器组. 前者称软栈，后者称硬栈. 软栈的一端是固定的，另一端是浮动的，浮动的一端称为栈顶. 所有信息的存入和取出都只能在栈顶进行，而且是在堆栈指示字寄存器的配合下，按 LIFO 原则工作的. 硬栈栈顶一端是固定的，压入的数据项占据栈顶，栈中原有的数据项都依次向栈底方向移动；弹出数据时，最后进栈的数据项先被弹出，下面各级数据项依次向栈顶方向移动.

软栈和硬栈功能均相同，主要作为 LIFO 缓冲存储器，用于中断处理、子程序嵌套及暂存数据. 在程序中断时，堆栈(自动)保存程序计数器、状态寄存器及工作寄存器的内容，并在中断结束后把这些现场内容恢复到相应的寄存器中.

堆栈(stack) 即“栈”.

存储系统(memory system) 由多种存储器、存储管理部件、存储管理软件所组成的系统. 当代计算机系统对存储系统的要求是存取时间短、容量大、价格低. 采用单一种类的存储器很难同时满足这三项要求. 采用存取时间短、价格高、容量小的高速缓冲存储器、主存储器和容量大、价格低、存取时间长的辅助存储器，在存储管理部件和软件的统一调度

管理下，组成一个多层次结构的存储系统，就能较好地满足这三方面的要求，使用户感到使用的存储系统具有高速缓冲存储器的工作速度、具有辅助存储器一样大的容量、而价格适中.

使用层次结构构成存储系统是依据中央处理器访问存储器的局部性，即在一段时间内，中央处理器存取程序或数据所访问的存储单元都集中在一个较小的连续地址区域内. 根据访问的局部性，可将当前集中访问的连续地址区的内容存放在高速缓冲存储器中，使中央处理器可通过高速缓冲存储器实现程序和数据的高速访问. 数据和程序在高速缓冲存储器和主存储器中的调动由硬件(存储管理部件)自动完成(参见“高速缓冲存储器”). 而主存储器与辅助存储器层次可用虚拟存储技术构成虚拟存储器，可为用户提供一个容量比主存储器大得多的可随机访问的可编程的逻辑地址空间. 程序和数据在主存储器和辅助存储器中的交换是在存储管理部件(硬件)和操作系统(软件)共同管理控制下自动实现的(参见“虚拟存储器”).

主存储器(main memory)　亦称内存储器. 是计算机的主要工作存储器. 用于存放 CPU 运行的程序和数据以及必要的系统软件，能由 CPU 直接随机进行存取的存储器，它也能被某些外部设备(如磁盘)直接访问. 主存储器是随机存取存储器，通常由磁芯存储器或半导体随机存储器组成，存取数据的时间与数据所在存储单元的地址无关. 其容量比高速缓冲存储器大，比辅助存储器小，而其工作速度则比高速缓冲存储器低，比辅助存储器高. 主存储器的主要指标有：

1. 存储容量，表示存储器可容纳的信息量，一般以字(W)、千字(kW)或字节(B)、千字节(kB)甚至用百万字节(MB)、十亿字节(GB)表示.

2. 存取周期，是允许存储器进行连续存取操作的最短时间间隔，一般以微秒(μs)或纳秒(ns)为单位.

3. 取数时间，存储器从接到读命令到数据到达数据缓冲寄存器并达到稳定所需时间.

内存储器(inner storage memory)　即“主存储器”.

辅助存储器(auxiliary memory)　一种存储器. 是存储系统中不能由中央处理器直接访问的大容量存储器. 辅助存储器存放的数据和程序不需先调入主存储器就能由中央处理器使用. 辅助存储器在存储系统中的主要作用是扩大存储系统的总容量. 辅助存储器中常用的存储设备主要是磁存储器和光存储器. 磁存储器主要有磁盘存储器和磁带存储器. 磁盘存储器存取速度快、存储容量大，是最常用的辅助存储器. 磁带存储器的存取速度比磁盘存储器低，联机存储容量比磁盘存储器容量小，但它常易于脱机存放，多用于存放需要长期保存的数据. 光存储器的存储容量大，存取速度接近于磁盘存储器，但它不易写入，常用于存放大量的程序和数据.

高速缓冲存储器(cache)　一种存储器. 是位于中央处理器和主存储器之间的高速小容量存储器. 高速缓冲存储器在高速工作的 CPU 和较慢速度的主存储器之间起速度缓冲的作用. 高速缓冲存储器的工作速度可与 CPU 相匹配，其容量通常比主存储器小很多，通常仅为主存储器容量的几百分之一，其内容是主存储器中的一个小存储区内容的复本. 高速缓冲存储器的内容可被 CPU 直接访问，也可与主存储器直接交换数据. 当 CPU 访问主存储器某单元时，若该单元的内容已在高速缓冲存储器中，则 CPU 仅需访问高速缓冲存储器，这种情况称为命中，命中时，访问时间由高速缓冲存储器的速度决定；反之，若该单元内容不在高速缓冲存储器中，则称为不命中，此时，CPU 应访问主存储器，并将包括该单元在内的一个小存储区调入高速缓冲存储器，以使下一次的访问能命中. 不命中时的访问时间决定于主存储器的访问时间. 目前，高速缓冲存储器的命中率达 90%以上，因此 CPU 对主存储器的 90%以上的可靠访问由高速缓冲存储器来实现，其访问时间接近于高速缓冲存储器的访问时间. 从而大大缩短了 CPU 获取指令和数据的时间，大大提高了处理机执行指令的速度.

高速缓冲存储器的工作机制体现了程序访问的局部性. 所谓访问的局部性是指在某一较短时间间隔内访问的地址仅集中在一个小的地址空间内. 高速缓冲存储器工作时命中与否的判定、主存储器地址与高速缓冲存储器地址的变换、替换策略及其实现、写一致性的实现等操作都是在硬件控制下自动完成的. 随着超大规模集成电路技术的发展，在高档微处理器芯片内都已集成了高速缓冲存储器，且有的将数据和指令分存在两个高速缓冲存储器中，它们分别称为数据高速缓冲存储器和指令高速缓冲存储器. 在此基础上，还可在微处理器与主存储器之间再设置第二级高速缓冲存储器，从而可使系统性能进一步提高.

联想存储器(associative memory)　亦称相联存储器. 一种存储器. 是按照给定内容特征访问，而不是按地址进行访问的一种存储器. 联想存储器工作时，将待查找的内容特征同联想存储器的所有存储单元的内容同时进行比较，若某单元内容与之相符合，则被选中，可弹出该单元的全部或部分数据，也可将新的数据写入该单元. 联想存储器的高速、并行操作能力和信息处理功能使之最适合用于信息检索和更新. 联想存储器由于电路结构复杂、成本高，

故一般存储容量较小.

相联存储器(associative memory)　即"联想存储器".

虚拟存储器(virtual memory)　由虚拟存储技术构成的存储器. 计算机系统中,在存储管理部件(硬件)和操作系统的统一管理下,在主存储器和辅助存储器层次上,为用户提供一个比主存容量大得多的可随机访问的可编程地址空间的技术. 这种虚拟存储技术构成的存储器称为虚拟存储器. 对用户来说,好像系统具有一个容量接近辅助存储器的主存储器,工作速度接近主存储器而价格又接近辅助存储器的存储器. 用户编程使用的虚拟存储器地址称为虚地址或逻辑地址. 程序运行时中央处理机实际访问的存储器仍然是主存储器,其地址称为实地址或物理地址. 用户编制的程序由操作系统装入辅助存储器,程序运行时和被访问的内容已在主存储器中,此时应将虚地址转换为实地址,因为中央处理机访问的是主存储器的实地址;如果访问的内容不在主存储器中,称为页面失效,操作系统将要访问的那一页从辅助存储器自动调入主存储器中,这个过程中虚实地址变换或替换策略的计算也是在存储管理部件和操作系统的统一管理下自动实现的. 存储管理部件负责虚实地址转换,操作系统负责调页管理、实存管理以及信息的自动调度.

虚拟存储器的管理方式分为页式管理、段式管理、段页式管理三种. 页式管理将逻辑地址和主存储器的实地址都分成固定大小的页;段式管理按程序的逻辑结构划分为段,段的大小不固定;段页式管理方式中,主存储器分为固定大小的页,而虚拟存储器则按程序逻辑结构分成大小不固定的段,每段再划分成若干与主存储器页面大小相等的页. 页式管理有利于存储空间的利用和调度,而段式管理则有利于程序的编译、执行、共享与保护,段页式管理则兼有二者的优点.

随机存储器(random access memory)　简称RAM. 一类存储器. 是可以通过指令、以相同速度随机地个别地访问每个存储单元、访问所需时间基本固定而与存储单元地址无关的一类存储器. 主要用于组成计算机主存储器等要求快速存储的系统. 按工作方式,随机存储器可分为静态随机存储器和动态随机存储器两类. 前者使用触发器存储信息,其特点是速度快,只要不断电信息就不会丢失,使用方便. 动态随机存储器利用 MOS 管极间电容储存的电荷来存储信息. 由于微弱漏电流的存在,使 MOS 管极间电容上存储的电荷会随时间而减少,直至漏完,存储的信息便丢失,因此动态随机存储器需要不断对 MOS 管极间电容上存储的电荷重新充电,使信息获得周期性再生,这种操作称为刷新. 但动态存储器功耗低,集成度高,价格低,应用更广泛.

只读存储器(read only memory)　简称 ROM. 一种存储器. 是其所存储的内容不能由计算机指令加以改变的存储器. 由于 ROM 所存数据一般是在装入机器之前事先写好的,比较稳定,不易改变,即使在断电后也不会改变;其次,ROM 的结构比较简单,读出又比较方便,因而常用于存储长期固定的数据或信息,如各种函数表、字符和固定程序等. 只读存储器所具有的存储数据功能,是由在制造过程中所用掩模决定的,所以也称为掩模只读存储器. 为了方便用户使用,也适合于工业化大批量生产,后来出现了可编程序只读存储器. 可编程序只读存储器允许用户自己编写并存储所需的数据,但由于它的任一单元都只能写一次,还是不太方便,为此又出现了可擦可编程序只读存储器.

外围设备(peripheral device)　计算机系统中除主机以外的全部附属设备. 它包括输入设备、输出设备、辅助存储器、终端设备、转换设备和数据准备装置等. 外围设备是计算机与外界交换信息的工具,其主要作用是:将外界的信息输入计算机;输出计算机要输出的信息;存储需要保存的信息;编辑、整理外界信息以便输入计算机等. 外围设备根据系统的作业需要和经济考虑,可以联机工作,也可以脱机工作. 外围设备发展趋向是:电子化;智能化;提高人-机通信能力;微型化;低噪声;高速度;高可靠性.

磁带存储器(magnetic tape memory)　一种存储器. 是以磁带为存储介质,由磁带机及其控制器组成的一种辅助存储器. 磁带的表面成分可分成 rFe_2O_3 磁带、CrO_2 磁带、合金磁带和复合磁带等;磁带机由磁带传动机构和磁头等组成,能驱动磁带相对磁头运动,用磁头进行电磁转换,在磁带上以顺序方式存取数据. 磁带控制器是中央处理器在磁带机上存取数据用的控制电路装置. 存储数据的磁带可脱机保存和互换读出. 磁带机按带速不同可分为高速机、中速机和低速机三种. 20 世纪 70 年代末,出现了数据流磁带机,其结构简单、体积小、成本低.

磁盘存储器(magnetic disk memory)　一种存储器. 是利用磁记录技术在稳速旋转的圆形磁性媒体表面进行数据存储的直接存取存储装置. 磁盘存储器由磁盘、磁盘驱动器和磁盘控制器组成. 磁盘是两面生成有可磁化的磁性介质的平面圆盘片. 根据磁盘片基体材料不同可分为硬盘和软盘两类,其盘片尺寸、记录密度均有多种不同规格. 根据安装方式磁盘可分为可换式和固定式两种,其中可换式磁盘的组件又有盘盒、盘组和头盘等几种类型. 磁盘驱动器是驱动磁盘高速稳定旋转,并驱动磁头按一定磁记录格式在磁盘表面逐位进行弹出或写入数据的装置. 磁盘驱动器又分为磁头臂移动型和固定型两类.

磁盘控制器是控制磁盘驱动器工作、完成数据的写入和读出的装置. 磁盘存储器将向着提高记录密度、提高存取速度的方向发展. 硬磁盘存储器在存取速度、存储容量、可靠性等方面均比软磁盘存储器高.

光存储器(optical memory)　一种存储器. 是利用激光技术来存取信息的存储装置. 其特点是存储容量大,可达 10^{10}bit 左右;存储密度高,可达 10^{5}bit/cm^{2};存取速度快;光束偏转无惯性、无磨损;成本低;易控制;可实现随机存取等. 它具有很大的发展前途. 光存储器按存储器方式可分为按位光存储器和全息照相存储器.

1. 按位光存储器. 每个二进制位以单独的聚焦点形式记录在存储介质上. 它用激光束改变存储介质的物理状态(如透射率或磁化状态等)进行信息的写入;在读出光束作用下,利用存储介质的不同物理状态对激光束的不同效应(如强度的改变或光束偏振面的变化等)来鉴别二进制信息的"1"或"0". 地址的选择是借助于激光束被偏转到存储介质上的不同位置来实现的.

2. 全息照相存储器. 以激光器为光源,利用全息照相的方法实现信息的存储. 信息以全息图的形式按页记录到存储介质上,信息的读出就是全息图的再现. 这种存储器是页式存储器,根据存储媒体的物理或化学特性的不同,可以做成只读存储器或可存取存储器. 它具有内部分级结构,可以过量存储,因此可靠性较高.

输入设备(input device)　输入信息的一种设备. 将待输入的各种形式的信息转换成计算机易于识别和处理的形式,并输入到计算机的设备. 计算机输入的信息有数字量、模拟量、字符、声音和图象等. 它们一般均能直接由计算机识别和处理,而不需经过相应的输入设备转换成离散的二进制编码形式. 常用的输入设备有低带输入机、卡片输入机、模-数转换器、光学字符阅读机、磁盘存储器、磁带存储器、光存储器、键盘、鼠标、光笔、触摸屏、跟踪球、控制杆以及新型的语言输入设备、文字输入设备、图象输入设备等.

输出设备(output device)　输出信息的一种设备. 将计算机处理获得的二进制编码信息转换成其他设备所需要的形式,或人们希望的形式并输出的设备. 常见的输出设备有各种打印机、显示器、绘图仪、磁盘存储器、磁带存储器、声音输出设备、图象输出设备等. 通过数-模转换器可将数字信息转换成模拟信息,并送到计算机控制系统实现过程控制.

转换设备(conversion device)　实现数字量与模拟量相互转换的设备. 它包括数-模转换器和模-数转换器. 是计算机用于工业控制领域的重要设备. 通常先利用各种传感器将被控对象的温度、压力、位移、转角、流量等物理量检测出来,并转换为连续变化的电流或电压(模拟量);再通过模-数转换器转换成二进制数字量送入计算机进行处理,产生形式为数字量的控制信息;再经过数-模转换器转换为连续变化的电流或电压的模拟量,控制相应的执行机构动作,实现温度、压力、位移、流量等被控对象的自动控制. 转换设备的指标主要有精度、分辨率、转换速度等.

终端设备(terminal device)　计算机的一种输入、输出设备. 经通信线路或数据传输线路向计算机输入、输出数据和程序的输入、输出设备,简称终端. 与计算机距离较远,需在通信线路上连结调制解调器的终端,称为远程终端;反之离计算机距离较近,不需要连结调制解调器的终端,称为本地终端. 终端按用途可分为通用终端和专用终端. 通用终端又分为交互式终端和远程批处理终端. 专用终端用于特定场合,例如飞机订票、旅馆客房预定、银行存取款等. 交互式终端又名分时终端或会话式终端,可使用户与计算机系统之间按分时操作实现人机交互作用或会话方式的数据处理. 电传打字机、键盘字符显示终端、汉字显示终端、图形显示终端、图象显示终端都是交互式终端. 远离计算机的用户通过远程批处理终端对远处的计算机进行批量输入、输出作业. 当前的因特网终端更是不拘一格、多种多样.

计算机体系结构(computer architecture)　系指软、硬件的系统结构. 有两方面的含义:一是从程序设计者的角度所见的系统结构,它是研究计算机体系的概念性结构和功能特性,关系到软件设计的特性;二是从硬件设计者的角度所见的系统结构,实际上是计算机体系的组成或实现(参见"计算机组织"),主要着眼于性能价格比的合理性. 亚当(Adam)等人为了说明和研究从程序设计角度所看到的计算机的属性(外特性),在 1964 年最先提出计算机系统结构的概念.

概念性结构与功能特性,这是从程序设计者角度所看到的计算机属性. 它包括机器内的数据表示、寻址方式以及对这些数据的运算和控制这些运算的执行等(即指令系统). 对于通用型机器,一般包括数据表示、寻址方式、寄存器定义、指令系统、中断机构、机器工作状态的定义和状态切换、机器级的输入、输出结构以及对信息保护的支持等.

计算机体系结构主要研究软件、硬件功能分配和对软件、硬件界面的确定. 20 世纪 70 年代以来,在计算机软件方面有了显著的进展. 计算机在性能、速度、价格、可靠性和组织、实现技术上虽比 20 世纪 50 年代末有了巨大的突破,但它们的系统结构却并没有什么明显的、突破性的进展. 绝大多数机器的系统结构仍然没有脱离冯·诺伊曼型的范围. 程序设

计者所见的系统结构和20世纪50年代末相比变化不大.例如,对于指令系统,程序设计者基本上仍然立足于20世纪50年代末的计算机系统结构观点来设计复杂得多的软件.传统计算机的硬件组成与高级语言和操作系统之间的严重脱节,给软件的可靠性、源程序编译效率以及系统的解题效率等方面带来不利的影响,这是计算机系统结构需要解决的重要课题.20世纪70年代出现的数据流计算机系统结构思想,把传统计算机的指令控制流控制方法改变为数据控制流的控制方法,从而有可能自动排除运算相关性的障碍,达到高度并行的目的.

计算机软、硬件功能分配,主要应从实现费用、对速度的影响和其他性能要求来考虑,亦即考虑如何分配能提高性能价格比.操作系统的基本的、通用的功能的硬化或固化,有利于提高操作系统的执行效率和速度,减少开销;而功能不稳定,即需要不断变化的,用软件实现有利于提供应有的灵活性.实现费用包括研制费用和重复生产费用.硬件的设计费用和重复生产费用都比软件大,宜于硬件实现的功能应该是稳定的、常用的、比较小的,而且是软件实现的速度下降会对计算机系统性能有较大影响的那些功能.硬件实现只是对产量大的计算机系统才有经济效益.

并行处理计算机系统(parallel processing computer system) 一种计算机系统.利用多个处理机或多个功能部件来同时执行多个任务或多条指令,或同时对多个数据项进行处理的计算机系统.并行处理计算机系统采用各种并行处理技术,以便大幅度地提高处理速度和解题能力.并行处理计算机主要指以下两种类型的计算机:能同时执行多条指令或同时处理多个数据项的单中央处理器计算机;多处理机系统.

第一类并行处理计算机,是在单处理机内广泛采用各种并行措施.这类计算机的结构主要有流水线方式、多功能部件方式、阵列方式和数据流方式等.第二类并行处理计算机,即多处理机系统,能提高系统的性能和可靠性.这类并行处理系统的结构特点主要反映在它的互联网络上,可用机间耦合度作为主要标志来加以描述.耦合度是系统的模块(处理部件)间互连的紧密程度,它是数据传输率、响应时间、并行处理能力等性能指标的反映,它主要取决于选取的互联拓扑和通信链路的类型.多处理机系统依其机间耦合度以及其操作系统之独立性和数目的多少,可大致分为三类:分布式系统、松耦合式系统及紧耦合式系统.

分布式系统(distributed system) 多处理机或多计算机系统.是一种结构和处理功能都分散的多处理机或多计算机系统,一般属于中等耦合系统.通常每台处理机或计算机均有其单独的操作系统,各自都可完成系统中指定的一部分功能或执行一部分处理任务,并共享系统资源.它通常是由一台大型计算机和若干台小型计算机或微型计算机组成.各处理机或各计算机分布在不同地点,通过互联网络连结在一起.这种系统可以是分布处理机系统,其中各处理机共享一主存储器;也可以是更为常见的分布计算机系统,其中每台计算机都有自己的存储器.它具有对本地事件响应迅速、节省数据传输、减轻主计算机的负担、简化系统程序设计等优点.

互联网络(interconnection network) 多计算机的互联硬件网络.由开关元件按照一定的拓扑结构和控制方式,将一个并行计算机系统中的多个处理机(结点)或多个功能部件、存储模块及各种外部设备相互连结起来,实现相互通信的硬件网络.互联网络可分为静态和动态两类:

1.静态互联网络中每一个开关元件和缓冲寄存器等分散、固定地与一个结点连结,以建立该结点与相邻结点间的被动连结通路,但程序运行时的连结关系不能动态设置.静态互联网络的主要优点是造价低,可伸缩性强.静态互联网络的拓扑结构常见的有环形、线形、树形、星形、立方体形、网格形和总线形等几种.静态互联网络常用于大规模并行计算机.

2.动态互联网络中的开关元件按程序运行的要求,通过控制信号控制互联网络中的开关元件建立起结点间主动可控的通信通路,因此,结点之间的连结可动态改变.动态互联网络又有交叉开关、交换开关和细胞开关等几种类型.动态互联网络相对于静态互联网络,造价较高,技术较复杂,但有较强的通用性,多用于多处理机系统.

计算机网络(computer network) 多台独立计算机互联的计算机系统.用通信线路把不同地点、不同用途的多台独立计算机互连起来,以实现资源共享和信息交换的系统.除了主计算机外,通常还包括用户通信接口、通信网络以及网络控制硬件和软件等.计算机网络能实现信息传输与信息处理功能的结合,使远距离用户能够共享网络中的硬件、软件和数据等资源,提高网络的资源利用率、系统可靠性和信息处理能力等.它已广泛应用于数据收集与交换、经营管理、武器控制、过程控制、信息服务、电子邮件、办公自动化等方面.计算机网络的结构形式有集中式、分布式和环形三种.其连结方法分直接连结和通过通信接口处理机间接连结两种.

网络协议(network protocol) 计算机网络中的规则和约定.计算机网络中各类实体(用户实体和资源实体)为了彼此通信、相互合作、完成任务,需要共同遵守的规则和约定.网络协议通常由语义、语法和变换规则三部分组成.

1. 语义. 规定了通信双方彼此之间准备"讲什么",即确定协议元素的类型.

2. 语法. 规定通信双方彼此之间"如何讲",即确定协议元素的格式.

3. 变换规则. 规定通信双方彼此之间的"应答关系",即确定通信过程中的状态变化,通常可以使用状态变化图来描述.

计算机网络大都按层次结构模型组织计算机网络协议,并按协议层次模型组织网络协议软件.

局域网(local area network) 一种计算机网络. 是通信范围较小的计算机网或数据通信网. 通常在数百米至数公里的范围内,物理信道总长度较短,使用低误码率物理信道,一般连结有几十至几百个属于同一组织的工作站(网络结点),数据传输率一般在 1MB/s 以上. 局域网的布局通常为总线形、环形、星形和树形等,连结比较方便,简单灵活,不占用邮电通信线,费用较低,安全保密性好,适用于一座办公楼、一个公司、一个机关或一所学校的管理、调度、控制和通信. 它能为用户提供的服务包括数据库和集中式文件存取、报文处理和电子邮件、批量和实时屏面信息传输等,适用于办公室自动化、分布式控制、分布式数据处理等环境.

城域网(metropolitan area network) 一种计算机通信网络. 在一个城市范围内(小于 100km 的范围),以光纤为主要传输媒体,将市内各局域网连结起来,实现高速率(100 MB/s 以上)的数据、声音和图象传输的通信网络. 它是城市内通信的主干网,并可向外接入广域网,实现城市间信息传输. 城域网的标准是 IEEE802. 6 分布式队列双总线(DQDB).

广域网(wide area network) 一种计算机通信网络. 是连结若干个城市、地区、国家间的通信网络. 其范围可达数千公里. 它们通常采用电路交换或包交换方式工作. 广域网的标准是以国际化标准组织(ISO)制定的开放系统互联参考模型(OSI/RM)以及 TCP/IP 为基础制定的.

因特网(Internet) 全世界最大的、开放的广域网. 它由众多的网络互连而成. 这些网络包括采用 TCP/IP 协议的网络(称为 IP 因特网)和一切采用开放协议的计算机网. 因特网是由美国的 ARPA 网发展演化而成的. ARPA 网是 1969 年美国国防研究计划局建立的. ARPA 网在 1983 年正式采用 TCP/IP 协议后,通过称为网关的网络互联设备实现了不同地区、不同国家的独立网络之间的互联(国际互联). 这种以 ARPA 网为中心的互联网被称为因特网. 1994 年以后,美国著名的校园网 NSF 网、欧洲的 EBONE 和 NORDUNET、英国的 PIPEX 和 JANET 以及日本的 WIDE 等都先后接入因特网. 随着后来因特网的商业化,全世界已有数万个网络、数百万台计算机接入了因特网,为因特网提供了极为丰富的信息资源和应用服务,使之成为世界上最大的互联网. 因特网的基本应用包括使用 FTP 协议的文件传送和匿名文件传送;基于 TCP/IP 的终端仿真协议的远程登录(Telnet);基于简单邮件传输协议 SMTP 的电子邮件;基于万维网(WWW)的多种服务.

客户机/服务器(client/server) 计算机体系结构中的一种新模式. 在局域网中,如果所连结的计算机在机器性能与逻辑地位上存在差别,其中功能强的机器在逻辑上处于"主机"位置,而功能弱的机器在逻辑上处于"从属"位置,前者称为服务器,后者称为客户机,该网络就形成客户机/服务器的体系结构. 这种结构已经战胜传统的大型计算机与小型计算机的结构,以更好的价格性能比为用户所接受. 如果服务器的功能远远大于客户机,就称为胖服务器/瘦客户机体系结构. 反之,也有瘦服务器/胖客户机的体系结构. 不同的体系结构有不同的物理配置,而且也有不同的程序设计方法. 这样的灵活性,为用户提供了广阔的选择余地. 这种体系结构不仅适应于局域网,而且适应于广域网. 在因特网上,就有众多的、各种类型的服务器,更有数量庞大的客户机.

通信(communication) 信息传输与交换的过程. 即信息从信源到信宿间的传输. 信源指信息的发送者(或设备);信宿指信息的接收者(或接收设备). 通信的方式通常分为有线通信和无线通信两种. 前者用电缆、光缆等作媒体;后者用电磁波、激光等作媒体. 在计算机系统中央处理机和外部设备之间、计算机网络之间的信息传输,均称为通信. 利用电磁系统传输信息,称为电信. 电信科学技术主要研究载有信息的电信号的产生、变换、处理、传输、交换和接收等过程的理论和技术,旨在研究、制造和改进各种技术系统,以保证在各种环境和距离的情况下提高通信的可靠性、有效性和经济性(参见本卷《通信与信息理论》同名条).

接口(interface) 计算机科学术语. 指两个不同系统(或设备)的交接部分. 在计算机系统中的接口一般可分为硬件接口和软件接口两类.

1. 硬件接口. 在计算机各部分(如中央处理机与通道、通道与外部设备控制器、中央处理机与主存储器等)之间、计算机与计算机之间、计算机与通信系统之间的连结设备. 它包括许多信息传输线及其逻辑控制电路. 例如,主机与外部设备相连结的输入输出标准接口,可使处理机的设计不受外部设备的种类和数量的限制,有利于提高系统配置的灵活性、可靠性和扩充性;另一方面也使控制器和外部设备的设计不受处理机型号的限制,有利于实现外部设备产品系列化和标准化.

2. 软件接口. 为系统与用户之间、子系统与子系统之间、两个程序块之间专门设计的程序. 软件支援环境要求软件接口都具有统一性. 例如,为了实现用户和各种软件工具的通信,要求有统一的调用方式.

计算机系统性能评价(computer system performance evaluation) 用以评估计算机性能的过程和方法. 指对计算机系统的各项性能指标进行测试、验证、分析和研究的过程和技术. 性能代表系统的使用价值. 性能评价和成本分析综合进行,可以获得各种系统性能和性能价格比的定量值. 其作用是指导新型计算机系统的设计和改进;指导计算机应用系统的设计和改进,包括选择计算机类型、型号和确定系统配置等.

计算机系统性能指标大致可分为可靠性指标和工作能力指标两大类. 可靠性指标主要包括可靠性、可用性和可维护性,此外还包括安全性、完整性等. 可靠性、可用性和可维护性三个指标统称为 RAS 性能. 可靠性、可用性、可维护性、完整性和安全性又统称为 RASIS 性能. RAS 性能和 RASIS 性能是在设计和生产计算机系统时必须考虑的. 这些性能的好坏直接关系到计算机的使用价值,是计算机用户比较关心的综合性指标. 工作能力指计算机系统在正常工作状态下所具有的能力. 常用的工作能力指标有系统生产率、利用率和响应特性等.

1. 生产率. 系统的生产率指在单位时间内系统处理的信息量. 系统生产率与计算机的字长和运算速度、主存储器容量和存取周期、通道信息流量的速率、输入输出设备配置等有关. 描述生产率的指标有指令执行速度、吞吐率等.

2. 利用率. 系统或其某一部分的利用率,指在一个评价期间内它的实际使用时间所占的比率.

3. 响应特性. 用系统从输入到产生响应之间的时间度量,通常包括响应时间、周转时间等.

性能评价技术主要有测量技术、模拟技术、分析技术等. 测量技术使性能成为数量化的、能进行度量评比的客观指标;模拟技术和分析技术可从系统本身或从系统模型获取有关的性能信息.

性能价格比(cost performance) 一种用来衡量计算机产品优劣的概括性指标. 性能主要指机器的运算速度、主存储器的容量和存取周期、通道信息流量的速率、输入输出设备配置 、机器的可靠性程度等. 价格则是指机器的售价. 如果一台计算机的性能价格比的值越高,就表明该机器越好.

可靠性(reliability) 评价计算机系统的一种指标. 在规定的工作时间内和工作条件下,一个系统、设备或元件能正确工作而不发生故障或失效的概率. 计算机系统的可靠性通常用平均故障间隔时间来表示;计算机器件的可靠性一般用单位时间内器件的失效几率来表示;计算机软件的可靠性是指软件系统在意料的情况下能够正确地工作(称为正确性),而在意外的情况下也能做出适当的处理以避免造成严重的损害(称为健壮性). 提高可靠性的措施主要有三条:

1. 对元器件进行严格筛选以提高其可靠性,即采用非容错法.

2. 采用各种容错技术和冗余技术.

3. 设计较完善的诊断程序和检测手段等.

可用性(availability) 亦称平均有效率. 评价计算机系统性能的一个指标. 指计算机系统的使用效率. 可用性以系统在任意时刻能正确运行的概率来表示

$$P=\frac{\mathrm{MTBF}}{\mathrm{MTBF}+\mathrm{MTRF}},$$

式中 MTBF 表示系统的平均无故障间隔时间,是系统能正确运行时间的平均值;MTRF 表示平均故障修复时间,是从发现故障到机器恢复正常工作的这段时间的平均值.

平均有效率(average effectiveness level) 即"可用性".

可维护性(maintainability) 用以描述计算机系统维修效率的一种指标. 通常用平均故障修复时间(MTRF)来表征. 故障修复时间指系统发生一次故障后使系统返回到正常状态所需的时间,它包括故障的诊断、定位、校正等所需的时间. MTRF 是系统发生故障后使系统返回到正常状态所需的平均时间. 可维护性反映出该系统所提供的诊断、检测和维修措施的完备程度. 系统的可维护性与系统中的软件可维护性密切相关. 软件可维护性指维护人员理解、改正、改动和改进这个软件的难易程度. 它与以下三个因素有关:

1. 可理解性. 表现为外来读者理解软件的结构、接口、功能和内部过程的难易程度.

2. 可测试性. 诊断和测试的容易程度,取决于软件可理解性、软件结构和可用的测试工具.

3. 可修改性. 软件容易修改的程度,与软件采用的设计原理和规则有关.

安全性(security) 评价计算机系统性能的一个指标. 它标志着程序和数据等信息的安全程度. 在现代社会中,计算机在社会事务管理、军事决策等方面的作用日益增大,必须解决信息的安全保密问题. 为了提高安全性,除要求用户方面建立完备的计算机使用管理制度和严格的保密措施外,系统应配备有防护能力的软件,提供信息盗窃和信息破坏的防护措施. 一般采用以下 5 种防护措施:

1. 限制法. 通过操作系统来限制不同用户对特定程度和数据的使用权限.

2. 验真法. 使计算机系统能判定用户的身份是否合法,并采取相应的措施.

3. 加密法. 对处理过程中的信息加密,使盗窃者即使取得数据,也很难破译.

4. 阻塞法. 计算机工作过程中发现可疑的调用指令时,立即采取措施以阻塞信息的输出和泄漏.

5. 侦察法. 跟踪和记录可疑的调用并发出警告信号,从而引起操作管理人员的注意,采取相应的防护措施.

完整性(integrity) 评价计算机系统性能的一个指标. 它标志着程序和数据等信息的一致性和完整性. 计算机配备文件系统之后,给用户带来了很大的方便,但一旦机器发生故障,或者由于操作人员的误操作而又未采取预防和补救措施,就可能使文件系统中的部分信息丢失或变动,从而损害整个系统信息的完整性和数据之间关系的一致性. 为保护信息的一致性和完整性,系统提供有完整性控制,用以防止错误操作或故障对所存储信息的影响. 系统还配备完备的文件管理系统或数据库管理系统,采用保留副本(拷贝)、定期卸出、断电保护等措施,以便一旦发生故障后仍能恢复,使损失尽量减小.

兼容性(compatibility) 评价计算机性能的一个指标. 指计算机设备或程序能兼用于多种不同系统的性能. 下面分别说明硬件兼容性和软件兼容性.

硬件兼容性指一种机器的目标码可以在另一种机器上执行,所得结果与原机器的执行结果完全相同. 实现硬件兼容性有两种途径:一是两种机器采用完全相同的指令系统;二是在第二种机器中设置一些专用指令,遇到该机器所没有的指令时,就将其翻译成对应的指令,使执行结果完全相同. 经过少许调整或不需调整就能用于其他系统中的部件、外部设备或其他设备,称为兼容硬件.

软件兼容性指用某种语言编写的源程序可在两种不同的机器上分别进行翻译并加以执行,所得的结果基本相同,但两种机器的指令系统可能有很大差异.

兼容性很少表现为完全互相兼容(完全互相兼容称为双向兼容). 一般为单向兼容,即一台机器的程序可移植到另一台机器上,反之,则不行. 单向兼容又可分为向上兼容和向下兼容两种. 前者是指老机器的程序一定能移植到新机器上使用;后者则指新机器的程序一定适用于老机器. 在同一计算机系列中,通常为向上兼容,即在较低档计算机上编写的程序可在较高档计算机上运行,而在较高档计算机上编写的程序往往不能在较低档计算机上运行. 计算机硬设备(如某插件、某外接部件等)与所连用的系统(如不同型号的计算机)之间也有兼容性问题,在概念上和软设备与系统的兼容性类似.

计算机软件

软件(software) 计算机系统的重要组成部分. 是计算机系统中规定并直接指挥计算机系统工作的程序及其文档的统称. 软件是计算机系统不可缺少的组成部分,由程序及其文档组成. 程序精确而完整地描述计算任务中的处理数据、处理动作和步骤;文档是对程序的解释和说明,也是使用、理解和维护软件所不可缺少的重要资料. 计算机的软件可以粗略地分为两层. 居于计算机系统中最靠近硬件的一层称为系统软件,使用频繁,但与具体应用领域无关;另一层则与具体应用领域有关,称为应用软件. 系统软件主要是对计算机系统进行调度、管理、监视及服务等,主要包括操作系统、各种语言处理程序、服务性程序、数据库管理程序、计算机网络软件以及各种标准程序库等. 应用软件的例子有过程控制程序、事务管理程序等,它们一般都通过系统软件发挥作用.

程序(program) 计算机软件的本体. 是为使计算机实现预期的目的而对其计算机任务中的处理对象和处理规则所进行的描述. 处理对象是数字、文字、图形、图象、声音等数据. 处理规则一般指处理动作和步骤. 作为一种具有逻辑结构的信息,其描述必须是精确的和完整的. 可以用低级语言(机器语言或汇编语言)或高级语言编写. 用低级语言写成的程序一般是一系列指令和有关的数据;用高级语言写成的程序一般是一组说明和语句. 程序要能实际起作用,必须装入到机器内部. 程序的实际工作过程称为程序的执行,程序的一次执行称为一个进程. 衡量程序质量主要考察以下两个特性:

1. 静态特性. 与执行过程无关的特性,衡量的标准主要是结构良好性,使之简明、易读、易维护.

2. 动态特性. 与执行过程有关的特性,衡量的标准有正确性、可靠性、坚固性、稳定性和适应性等.

源程序(source program) 一种计算机程序. 是用不同于机器语言的任何一种源语言编写的计算机程序. 由于汇编语言和高级语言都经过了规范化,因而源程序要经过某种方式的翻译后才能转变成机器能够直接执行的用机器语言表示的"目标程序";如果源程序使用的源语言是汇编语言,则称这种翻译过程为"汇编";如果源语言是 FORTRAN, PASCAL, COBOL 等编译性高级语言,则称这个翻译过程为"编译";如果源语言是解释性高级语言,如 BASIC 等,则称这个翻译过程为"解释".

主程序(main program) 一种计算机程序. 是能调用子程序而又不被任何子程序所调用的程序. 一个程序只能有一个主程序,并且程序由此开始执

行.主程序包括程序语句、说明语句、可执行语句、结束语句等.说明语句定义变量和常数,在程序中它们必须在可执行语句之前.可执行语句可以产生一条或几条可执行的机器指令.

子程序(subprogram) 一种计算机程序.是程序中具有相对独立性的程序单位.子程序可以独立编译.当一个程序中有多处均需要执行某一特定功能时,可把这个特定功能当做子程序处理,只编写一次,而在需要的不同点使用.控制主程序进入子程序的过程称为子程序调用.从子程序回到主程序的过程称为子程序返回.子程序能自动返回调用它的程序.子程序还可调用其他子程序,这就是子程序嵌套.子程序还可以直接或间接地调用子程序自身.允许自己调用自己的子程序称为递归子程序.

例行程序(routine) 亦称例程.一种计算机程序.是与一项计算任务相对应的处理对象和处理规则的描述.可以是一个主程序的一部分或一个专用程序,也可包含若干个子程序.它一般在一个程序或多个程序中多次使用.例行程序和子程序常常存放在程序库中,通常存储在外存储器上.

例程(routine) 即"例行程序".

子例行程序(subroutine) 一种计算机程序.是与子计算任务相对应的处理对象和处理规则的描述.例如,计算初等函数、三角函数、对数函数、二进制到十进制的数制转换、十进制到二进制的数制转换等,都可编为子例行程序.子例行程序有两个方面:一个是定义方面,称为子例行程序定义或子例行程序说明;另一个是调用方面,称为子例行程序调用.它一般是多次使用,通常单独编写.使用时,主程序通过子例行程序调用指令将控制转向子例行程序,执行完毕之后再自动返回原来的主程序.

翻译程序(translator) 一种实用程序.把一种程序设计语言(称为源语言)编写的程序(称为源程序)翻译成与之等价的另一种程序设计语言(称为目标语言)所表示的程序(称为目标程序)的软件.翻译程序有两种典型的实现途径:

1.生成性途径.先由翻译程序把源程序静态地翻译为目标程序(通常是机器语言程序),然后再由计算机执行目标程序,并利用输入数据产生所需的输出结果.生成性途径能翻译出运行效率较高的目标程序,但灵活性较差,且算法较繁,不易实现.它主要适用于以加工和计算为主、要求高效运行的场合,通常用于实现面向过程的程序设计语言(例如FORTRAN,ALGOL,COBOL,PL/1,PASCAL)的翻译程序,一般在成批处理方式下工作.

2.解释性途径.按照源程序中语句的动态执行顺序逐句翻译,并立即予以执行.解释性途径易于实现,比较灵活,但重复执行的语句要多次重复翻译,运行效率较低.它适用于以研究、调试和修改程序为主的场合,通常用于实现各种交互会话式语言(例如BASIC,APL)的翻译程序.

翻译程序按其实现途径可分为生成性翻译程序和解释性翻译程序.后者简称解释程序.前者常用的有编译程序、汇编程序、交叉编译程序、交叉汇编程序、反编译程序和反汇编程序等.

汇编程序(assembler) 一种翻译程序.把汇编语言书写的程序翻译成与之等价的机器语言程序的翻译程序.汇编程序输入的是用汇编语言书写的源程序,输出的是用机器语言表示的目标程序.汇编程序分为简单汇编程序、模块汇编程序、条件汇编程序、宏汇编程序和高级汇编程序等.

1.简单汇编程序.汇编后的机器语言程序直接放在内存之中准备执行.由于简便而得到广泛使用.

2.模块汇编程序.为适应模块程序设计方法而研制的一种汇编程序.

3.条件汇编程序.满足一定的条件时才进行汇编的程序.适用于编写选择性较大的程序或程序包.

4.宏汇编程序.能实现具有宏加工功能的汇编程序.

5.高级汇编程序.采用高级语言的控制语句结构的汇编程序.

解释程序(interpreter) 一种实用程序.按照源程序的指令或语句的动态执行顺序,逐条或逐句翻译,并立即执行的翻译程序.解释程序的特点是把源程序直接翻译成加工输入数据,产生所求的操作,而不是产生目标语言形式的目标程序.解释程序的突出优点是可简单地实现,且易于在解释执行过程中灵活、方便地插入修改和调试措施,但最大缺点是执行效率很低.解释程序包括两大部分:第一部分是"解释循环程序"和"控制循环程序";第二部分是一组例行子程序,每个子程序能够执行源程序的一种语句.

编译程序(compiler) 一种实用程序.把用高级程序设计语言书写的源程序,翻译成等价的计算机汇编语言或机器语言的目标程序的翻译程序.一个实用的编译系统,除了把源程序翻译成目标程序这个基本功能外,还应具备语法检查、调试措施、修改手段、覆盖处理、目标程序优化、不同语言合用和人机联系等重要功能.早期编译程序的实现方案,是把这些功能完全收纳在编译程序之中.目前习惯做法是在操作系统的支持下,配置调试程序、编辑程序和连结装配程序.编译过程可分为分析和综合两个阶段.编译程序在分析阶段建立一些表格,改造源程序为中间语言形式,这些工作是经过词法分析、语法分析和语义分析三个逻辑步骤实现的.编译程序在综合阶段经过代码优化、存储分配和代码生成等三

个逻辑步骤，根据符号表和中间语言程序产生出目标程序. 随着程序设计语言在形式化、结构化、直观化和智能化等方面的发展，作为实现相应语言功能的编译程序，也正向自动程序设计的目标发展，以便提供理想的程序设计工具.

交叉翻译程序(cross translator) 一种实用程序. 在某台计算机上运行，产生另一台计算机的机器语言程序的翻译程序. 当运行目标程序的计算机内存较小，难以支持翻译程序运行(如单板机)，或其程序研制环境较差不易研制系统软件或应用软件(如向量计算机)时，可采用交叉翻译的办法，即在一台研制环境较好的计算机上，配置一个能产生目标机的机器语言程序的翻译程序(编译程序或汇编程序)，这种翻译程序称为交叉翻译程序(交叉编译程序或交叉汇编程序).

逆翻译程序(inverse translator) 亦称反翻译程序. 一种计算机程序. 与通常的翻译程序的翻译过程相反，它把某一种机器语言的程序翻译成与之等价的另一种高级语言或汇编语言的程序的软件. 逆翻译程序(逆编译程序或逆汇编程序)通常在剖析只有目标程序的软件的内部结构和实现算法时，或是在调试目标程序需要给出源程序一级上的调试信息时使用.

实用程序(utility program) 系统软件中的一组常用程序. 它向操作系统、应用软件和用户提供经常需要的功能，如编辑程序、连结装配程序、文件操作程序和调试程序等.

引导程序(bootstrap) 一种实用程序. 计算机启动后，系统自行输入的一个简短的程序，由它来完成引导过程. 引导是计算机开始输入时常采用的一种手段. 计算机启动后，由设备来实现少量指令和数据的输入，然后由它们再输入其他程序，这种过程称为引导. 微型计算机中的引导程序可将用户通过键盘打入的程序及数据引导输入到随机存储器中.

编辑程序(editor) 一种修改和产生文件的程序. 用户利用编辑程序对存储在计算机中的文件进行增加、删除、修改和剪贴等加工处理. 依据编辑文件内容的不同，编辑程序可分为两类：第一类是正文编辑程序，编辑文件的内容是正文. 正文泛指文稿、程序、结构化框图等对象的字符和文字表示. 第二类是图形编辑程序，编辑文件的内容是图形信息. 采用多窗口编辑技术后，编辑程序能够把一个显示屏幕划分成若干个不相交或相交的区域. 例如正文、框图、命令、图形等多种编辑状态，可被分配在这些区域里并操作.

连结装入程序(linking loader) 一种实用程序. 它把多个经过编译或汇编的目标模块连结装入成一个完整的可执行程序. 连结装入程序由连结编辑程序和再定位装入程序组合而成. 连结编辑程序是对程序库进行扫描，将整个程序涉及的目标模块逐个调入内存并连结在一起. 再定位装入程序能把目标模块的相对地址加工成绝对地址. 连结装入程序不仅为个别编译提供了连结装入能力，而且还为用户直接调用程序库中的程序带来方便.

诊断程序(diagnostic program) 一种实用程序. 专门用于检测和定位计算机硬件设备故障或软件错误. 它是专供计算机应用人员或系统管理程序使用的程序. 例如，计算机应用人员需检测硬件设备的故障时，可运行相应的诊断程序，使计算机单步操作，由此跟踪各逻辑部件的逻辑状态，各逻辑状态可在程序支持下显示或打印出来. 有用于程序设计排错的诊断程序和用于数据检错的诊断程序；有用于检查整个系统故障的系统诊断程序；还有一种装在随机存取存储器中的 RAM 诊断程序，它利用联机实时仿真器对用户处理机的全部状态提供面板显示能力，因而能迅速查出各种问题. 有的诊断程序既可用于对硬件设备故障检测，又可用于对程序的错误定位.

调试程序(debugging routine) 一种实用程序. 是为检测并排除程序中的错误而编制的一种常用程序. 调试程序可用来帮助程序的研制和调试. 为被调试的程序设置和清除断点；跟踪程序的执行(即在逐步执行程序时打印出存储器中的必要的地址状态，以掌握运行状态)；显示和修改指定的寄存器或存储单元的内容；打印指定存储单元的内容；在存储器和磁盘，或存储器与打印机之间进行存储器转储等.

程序设计(programming) 设计、编制和调试计算机程序的过程. 在用计算机解决问题时，必须事先详细地安排好完整的解题计划，并以程序的形式把问题的数据、解决问题的方法和步骤都送入机器，机器才能按人的意图进行工作. 用计算机所能接受的形式把解决问题的计划用机器指令或程序设计语言等描述出来，这就是程序设计的过程.

程序设计的发展可归结为：从顺序程序设计到并发程序设计、分布程序设计；从非结构程序设计到结构程序设计；从低级语言工具到高级语言工具；以具体方法到方法学等.

结构程序设计(structured programming) 简称 SP 方法. 一种设计程序的技术. 采用自顶向下逐步求精的设计方法和单入口单出口的控制结构. 结构程序设计的概念由迪克斯特拉(Dijkstra，E. W.)于 1965 年首先提出，并经伯姆(Böhm，C.)和贾克比尼(Jacopini，G.)于 1966 年证明，指出任何一个正确程序可仅由顺序、选择和循环三种基本控制结构组成，同时尽量避免使用转移语句. 有时，人们将

使用三种基本结构、自顶向下逐步求精、主程序员组等一系列的技术，这些技术统称为结构程序设计.在软件工程中采用这种技术，可以收到较好的效果.在软件工程的初步设计阶段，采用自顶向下逐步求精的方法，可以把一个复杂问题的解法分解细化成一个由许多模块组成的层次结构的软件系统；在详细设计或编码阶段，采用自顶向下逐步求精的方法，可以把一个模块的功能逐步分解细化为一系列具体的处理步骤或某种高级语言的语句.使用结构程序设计技术有以下主要好处：

1.自顶向下逐步求精方法符合人们解决复杂问题的普遍规律，因此可显著提高软件开发工程的成功率和生产率.

2.用先全局后局部、先整体后细节、先抽象后具体的逐步求精过程开发出的程序有清晰的层次结构，容易阅读和理解.

3.不使用转移语句，仅使用单入口和单出口的控制结构，使程序的静态结构和它的动态执行情况趋于一致，因而使程序易于理解和验证.

4.程序的逻辑结构清晰，有利于程序正确性证明，有利于测试和维护等.

并发程序设计(concurrent programming) 一种程序设计.是设计、编制和调试并发程序的技术.一个并发程序由若干个可同时执行的程序模块组成.这种可同时执行的程序模块称为并发进程，由数据和有关的语句序列组成.并发进程的执行在时间上是重叠的，因此，一个并发程序的多个并发进程可以同时在多台处理器上并行执行，也可以在一台处理器上穿插执行.采用并发程序设计方法的好处是：可以使计算机的外围设备和处理器并行工作，缩短程序执行时间，提高计算机系统的效率.要有效地采用并发程序设计，必须提供并发程序设计语言，防止死锁的算法和同步机制等.

1.并发程序设计语言.并发程序设计语言的主要特征是引入了进程概念，这种语言提供实现进程同步和通信的手段.这类语言有并发 PASCAL，Modula-2，ADA 等.

2.死锁的预防和检测.进程因争夺资源而无休止地相互等待称为死锁.解决死锁问题有两种途径：一是预防死锁，设计各种资源调度算法，积极地防止死锁的发生；二是检测死锁，当死锁发生时能及时地发现并进行排除.

3.同步机制.在并发程序设计中，各进程对公共变量的访问必须加以制约，这种制约称为同步.进程的同步是通过同步机制实现的.代表性的同步机制有 PV 操作和管程等.PV 操作是最早提出的用来解决公共变量问题的同步操作.管程是指一组公共数据同与其有关的操作的集合，只有引用管程中的操作才能访问管程中的数据.

分布式程序设计(distributed programming) 一种程序设计.是设计、编制和调试分布式程序的技术.一个分布式程序由若干个可独立执行的程序模块组成.这些程序模块分布于一个分布式计算机系统的几台计算机上，并同时协调执行，它们在执行中需要交换数据，即通信.只有通过通信，各程序模块才能协调地完成一个共同的计算任务.对分布式程序的要求应该是：适应基于信件传递的同步机构，考虑实际的并行进程，减少进程间的通信开销等.

采用分布式程序设计方法解决计算问题时，必须提供分布式程序设计语言和相应的分布式算法.分布式程序设计语言应能有效地描述分布程序，说明进程及其在各处理机上的分布、进程通信等.它往往可以由一种常用增加分布和通信功能的程序设计语言而构成.

函数式程序设计(functional programming) 简称 FP.一种程序设计.是设计、编制和调试函数式程序的技术.函数式程序是由一些原始函数、定义函数和函数型组成的函数表达式.函数式程序具有引用透明性，即函数值惟一地由其参数值所确定，只要使用相同的参数值，对此程序的不同的调用总是得到相同的结果.引用透明性有利于程序的模块化.函数式程序设计的概念是由美国计算机科学家巴科斯(Backus，J.)于 1977 年首先提出的.巴科斯提出的函数式程序设计是引用函数型产生新函数，程序设计时从一般的对象空间上升到函数空间，因而具有优越的数学性质，有助于程序的理解、推理和验证.

逻辑型程序设计(logic programming) 一种程序设计.是设计、编制和调试逻辑型程序的技术.逻辑型程序设计是由柯娃尔斯基(Kowalski，R.A.)及考莫瑞尔(Colmerauer，A.)首先提出的，1972 年，他们提出"逻辑可当做程序设计语言来使用".柯娃尔斯基提出"算法＝逻辑＋控制".这里的逻辑是指对所解问题的描述，即程序的说明性语义；而控制则是如何解题的描述，即程序的过程性操作语义.理想的逻辑型程序设计只需程序员说明一个算法的逻辑部分，而控制部分完全应由系统来解决.逻辑型程序包括一组以一阶谓词逻辑的 horn 子句组成的规则，其每一规则可以翻译成 IF-THEN 形式的逻辑语言，可以翻译成能由解释程序执行的过程.计算则是推理过程的结果.典型的逻辑型程序语言是 PROLOG.针对 PROLOG 低效的回溯型控制策略，已经提出逻辑型程序的并行解释模型和并行逻辑程序设计.逻辑型程序设计由于在表达知识和建立软件的快速原型示范方面的明显优势，已经应用至自然语言处理、数据库、专家系统、符号代数处理、程序变换和验证、面向对象程序设计等领域.

面向对象程序设计(object-oriented programming)　一类程序设计方法.是利用面向对象程序设计语言进行设计、编制和调试程序的技术.实质上是一门代码组装技术,即讨论代码提供者(系统程序员)如何将信息以及对信息的处理进行封装,生成对象、类以及子类等,再提供给代码使用者(应用程序员)使用;代码使用者如何利用已有的类以及通信与继承规则生成新的类、子类与实例,从而生成所需要的应用程序.对象所存储的信息是被封装的,其内部变化不影响到系统的其他部分,从而使信息得到严密的保护,编出的程序易于调试与维护.类之间有继承关系,可将类中的方法实现自动地传播给他所有的子类,形成一个可重用的代码库,从而使编程容易而不会出错.

程序设计语言(programming language)　一类计算机语言.是用于书写计算机程序的语言.程序设计语言有三个方面的因素:语法、语义和语用.语法表示程序的结构或形式;语义表示程序的含义;语用表示程序与使用者的关系.程序设计语言的种类很多,但其基本成分不外四种,即数据成分、运算成分、控制成分和传输成分.程序设计语言的分类有多种方式:

1.按语言级别分为低级语言和高级语言.

2.按用户要求分为过程式语言和非过程式语言.

3.按应用范围分为通用语言和专用语言.

4.按使用方式分为交互式语言和非交互式语言.

5.按成分性质分为顺序语言、并发语言和分布语言.

6.按设计背景分为冯·诺伊曼式语言和非冯·诺伊曼式语言等.

面向机器语言(machine-oriented language)　一类程序设计语言.是围绕特定的计算机或一类计算机而设计的程序设计语言.这种语言符号形式表示机器指令,用助记符代替机器的操作码,用标识符代替地址码和变址码.其目的是让程序员摆脱计算机的一些纯事务性的细节问题(如不必硬记机器指令代码,摆脱2～10进制数转换和存储分配等问题),从而集中精力去考虑程序中的内在联系.这类语言的代表是汇编语言(又称符号机器语言).

面向过程语言(procedure-oriented language)　一类程序设计语言.是为说明处理过程而设计的一种与机器无关的程序设计语言.使用这种语言在计算机上解题时,人们不必去了解计算机的内部逻辑,只需集中精力考虑解题方法、算法的逻辑和过程的描述.由于这种语言对解题过程的描述采用了比较接近人们习惯的方式,因而易学、易懂、易使用,是用户广泛使用的一类程序设计语言.这类语言的代表有ALGOL,FORTRAN,COBOL,PL/1,PASCAL和ADA等.

非过程语言(nonprocedural language)　一种计算机程序设计语言.它同过程语言的区别是非过程语言编写的程序可以不必遵循计算机执行的实际步骤.在编写非过程语言程序时一般要遵守下列步骤:

1.执行内务处理,例如清存储器、预置计数器初值等.

2.打开文卷,检查标号.

3.读输入记录.

4.执行算术或逻辑运算.

5.产生输出记录.

6.重复步骤3～5,直到处理完全部记录.

7.打印最终的输出记录,关闭文卷.

8.结束程序或将控制返回操作系统.

然而,在非过程语言中,只要指明输入记录、所要完成的加工以及输出形式,便能得到所要求的输出结果,其余工作全部由系统来完成.其主要优点是可减少程序员编程的工作量.

当前使用的非过程语言种类很多:一类是以LISP,FP等为代表的函数式程序设计语言;一类是以PROLOG为代表的逻辑型程序设计语言;还有一类称为面向对象的程序设计语言,其典型语言是Smalltalk.

面向问题语言(problem-oriented language)　一类程序设计语言.是为了易于定义和解决某些问题而设计的一类与机器无关的程序设计语言.使用这类语言在计算机上解题时,人们不仅不需要了解计算机的内部逻辑,而且也不必关心问题的解法和计算过程的描述.只要给出问题的描述、输入数据和要完成的加工及输出形式,便能得到所要求的结果.其主要优点是减少了程序员编写程序的工作量.例如,报表语言、判定语言、机床控制专用语言、医学诊断专用语言、电路设计专用语言等,都是面向问题的语言.

系统程序设计语言(system programming language)　一类程序设计语言.是用来书写系统程序(如编译程序和操作系统等)的程序设计语言.一般可分为通用系统程序设计语言和专用系统程序设计语言两类.前者指可以针对各种不同的系统书写其系统程序的语言,亦即它是与系统无关的.例如PASCAL,并发PASCAL,MODULA,XCY,ADA等语言都是通用系统程序设计语言.后者是与系统有关的,它们只适宜书写特定系统的系统程序.例如,BLISS语言是专用系统程序设计语言,它适于为PDP-10机器书写系统程序.系统程序设计语言可

以缩短系统程序的研制周期，提高其生产率．随着开发性和分布性研究的深入，系统程序设计语言除作为书写系统程序的工具外，还将是整个计算机系统设计中的重要依据，是整个计算机系统的重要组成部分．

函数式语言(functional language)　一类程序设计语言．是一种非冯・诺伊曼式的程序设计语言．函数式语言主要成分是原始函数、定义函数和函数型．这种语言具有较强的组织数据结构的能力，可以把某一数据结构(如数组)作为单一值处理；可以把函数作为参数，其结果也可为函数，这种定义的函数称为高阶函数．程序就是函数，程序作用在结构型数据上，产生结构型结果，从根本上改变了冯・诺伊曼式语言的“逐词”工作方式．

函数式语言具有简明性和独特的表达能力，因此可用它来研究传统程序设计语言的语义．巴科斯(Backus，J.)分析了传统程序设计语言的缺陷，认为这些缺陷主要是采用冯・诺伊曼式系统结构所造成的．巴科斯提出的函数式程序设计系统 FP 摆脱了传统的冯・诺伊曼计算机结构，因此，需要一种新型的非冯・诺伊曼式的系统结构为后援．

逻辑型语言(logic language)　支持用逻辑型程序设计(参见“逻辑型程序设计”)方法求解问题的语言工具．典型的逻辑型语言是 PROLOG．它以一阶谓词逻辑的 horn 子句为语法，以 Robinson 归结原理为语义，加上深度优先的控制策略而形成的通用语言．它有内设的形式匹配和回溯功能．由于它出于运行效率的考虑，附加了控制机构，因而破坏了完备性，故不能看成是纯粹的逻辑型语言．逻辑型语言的一个发展方向是并行的逻辑型语言，如英国 ALVEY 计划的 PARLOG、并发 PROLOG 等．另一发展方向是将逻辑型语言和函数型、面向对象语言相结合或渗透，构造出新的语言，如 LOGLISP，POPLOG，FUNLOG，LOOPS 及 IDEAL 等．

面向对象语言(object-oriented language)　一种程序设计语言．面向对象语言是在抽象数据类型和模块化程序设计的研究基础上发展起来的一种程序设计语言，它的典型代表是美国施乐公司 PARC 研究中心开发的 Smalltalk-80 语言．它将程序系统看做是一组互相传递消息的活动着的对象，程序的说明性知识和过程性知识都装入在各个对象中，从而实现高度模块化．如果一组对象具有相同的私有存储描述和相同的操作集合，则可称为类．每个对象就是这种类的一个实例．面向对象的程序设计就是创建新的类和新的实例、创建子类、规定新的操作等．子类具有继承父类特性的能力．面向对象的程序设计具有很多优点：对象所存储的信息是被保护的；对象是模块化的；用户可以存取已有的对象，也可以创建新对象或修改原有对象；对象间的通信可以通过消息传递进行，因而易于用户观察和对话．一些新的面向对象的语言已经开发，如 C＋＋objective C，JAVA 等．面向对象的特点已广泛受到重视．

低级语言(low level language)　一类程序设计语言．指字位码、机器语言、汇编语言一类的程序设计语言．低级语言的特点是与特定的机器有关，功效高，但使用复杂、繁琐、费时、易出差错．其中，字位码是计算机惟一可直接理解的语言，但由于它是一连串的字位，因此较复杂、繁琐、冗长，几乎无人直接使用．

机器语言(machine language)　一类计算机语言．指不经翻译即可为机器直接理解和接受的程序语言或指令代码．机器语言使用绝对地址和绝对操作码．不同的计算机都有各自的机器语言，即指令系统．从使用的角度看，机器语言是最低级的语言．

汇编语言(assembly language)　一种面向机器的程序设计语言．是用符号表示指令操作码和地址的一种低级程序语言．汇编语言通常是为特定的计算机或计算机系列专门设计的，与机器语言很接近．一般的汇编语言中，一个语句与一条机器指令相对应．但有的汇编语言中可以有宏指令，它与一串特定的机器指令相对应，这样的汇编语言有时也称为宏汇编语言．这种宏汇编语言的结构级比汇编语言高，但比高级语言低．

高级语言(high level language)　一类程序设计语言．在一定程度上与具体机器无关，是一种易学、易用、易理解、易维护的程序语言．高级语言的表示方法要比低级语言更接近于待解问题的表示方法．用高级语言编写的源程序必须经过编译程序或解释程序翻译成目标程序才能被计算机执行．当高级语言程序翻译成相应的低级语言程序时，通常一条高级语言语句要对应多条机器指令，相应的目标程序往往功效较低．

常用的高级语言有 FORTRAN，BASIC，COBOL，PASCAL 和 PL/1 等．

数据(data)　计算机科学的基本概念之一．是可由人工或自动化手段加以处理的数字、文字、图形、图象、声音等符号的集合，是用于描述对象、概念、状态或情况的信息基本单元．数据是信息的存在形式和表示形式．信息以数据的形式被处理、使用和传播．计算机数据处理主要包括数据的采集、转换、分组、组织、计算、存储、排序、检索和传送等操作．数据经过处理解释过程形成信息．计算机系统的各种操作就是对数据进行各种处理，所以数据和程序都是软件的基本对象．

数据类型(data type)　数据的抽象结构的说明形式．同一类型的数据，其各成分都具有相同的特

性，可进行相同的操作.程序中出现的每个变量、常量或表达式的值，都属于确定的数据类型.变量的所有可能取值，以及在这些值上可允许的操作，都由数据的类型在程序中显式地或隐式地规定.数据类型可分为简单类型和复合类型两种：

1.简单类型.具有相同的数学特性和相同的计算机内部表示法，其数据的逻辑结构特点是只包含一个初等项的结点.通常有五种基本的简单类型：整数类型、实数类型、布尔类型、字符类型和指针类型.

2.复合类型，或称组合类型或结构类型.是由简单类型用某种方式组合而成的.根据不同的构造方法，可构成以下不同的数据结构类型：

1）数组类型.所有成分都属于同一类型.

2）记录类型.各成分不一定属于同一类型.

3）集合类型.它定义的值集合是其基类型的幂集，也就是基类型的值域的所有子集的集合.

4）文件类型.属于同一类型的各成分的一个序列，这个序列规定各成分的自然次序.

5）递归数据结构.在数据结构的定义式中出现名本身的数据结构.

抽象数据类型(abstract data type)　一种重要的数据类型.它是由一组对象和一组在这些对象上的操作所组成，利用数据抽象机制把一个数据类型的表示和操作汇集起来，封装在一起，定义成一个新的数据类型，这种新的数据类型统称为抽象数据类型.抽象数据类型不计实现的细节，对象的行为可以用所定义的操作完全予以规定.抽象数据类型在程序中是一个封闭的单元，它包含的信息有：

1.外部可见的：类型名、操作名、程序单元的头标，还可以给出类型变量值的规范形式和操作的特性.

2.外部不可见的：内部所用的类型、程序的主体，以及只能在内部调用的隐蔽程序.

数据结构(data structure)　由简单类型的数据构造复合类型数据的方法和表示.计算机软件中的数据结构一般包括数据的逻辑结构、数据的物理结构和数据的运算等三个方面.数据的逻辑结构描述数据间的逻辑关系，可以用一个二元组 $B=(K,R)$ 来表示.其中 K 是结点的有穷集合，结点(或称元素)是数据结构中讨论的基本单位；R 是 K 上的关系的有穷集合.数据结构分为线性结构和非线性结构.

1.线性结构.有且仅有一个终端结点和一个开始结点，并且所有的结点都最多只有一个前驱和一个后继.向量、栈、队列等顺序表以及字符串、链表等线性表都是线性结构.

2.非线性结构.树形结构、图、多维数组、稀疏矩阵、广义表等都是非线性结构.

数据的物理结构即数据的存储结构，描述数据的逻辑结构在计算机存储器的表示方式.通常有四种基本的存储映像方法：

1.顺序的方法.把逻辑上相邻的结点存储在物理上相邻的存储单元里，结点之间的关系由存储单元的邻接关系来体现.这种方法主要用于线性的数据结构.对非线性结构也可以采用局部线性化的方法实现顺序存储.例如，在树形结构中可以把结点按某种规则排成序列，用顺序存储方法把结点内部的信息稠密地存放在一起，而对结点之间的关系采用其他的存放方法.

2.链接的方法.将结点所占的存储单元分为两部分，分别存放数据项和指针项.

3.索引的方法.用结点的索引号来确定结点的存储地址.

4.散列的方法.在结点 k 的字段里取一个或几个字段的值 W_{ik} 作为关键码，结点 k 对应的存储地址 $\mathrm{LOC}(k)$ 由函数 f(称为散列函数)确定，$\mathrm{LOC}(k)=f(W_{ik})$.

数据的运算是定义在数据的逻辑结构上的，但运算的具体实现要在物理结构上进行.数据的运算包括对结点进行的检索、插入、删除、更新、排序等.数据结构尚有静态和动态之分.静态结构就是数据的结构(逻辑结构和物理结构)特性在该数据结构存在期间是不能改变的，例如向量、数组、记录等；而动态结构是在整个使用期间，数据的结构特性是可变化的，例如栈、队列、链表、树、动态数组结构、递归数据结构等.

数据模型(data model)　数据库系统中用以提供信息表示和操作手段的形式构架.它包括数据库数据的结构、操作和完整性规则等三部分.按照数据结构的不同，数据模型可分为层次模型、网状模型和关系模型三种.前两种采用格式化的结构，其数据结构与图相对应(图的顶点代表实体，用记录型表示；顶点间的连结弧表示记录型之间的联系).对应于树形图的数据模型为层次模型；对应于网状图的数据模型为网状模型.关系模型为非格式化的结构，用单一的二维表的结构表示实体及实体之间的联系.格式化模型可以描述复杂的数据结构，用存取路径实现数据间的联系；但程序和数据的独立性低，影响应用程序员的工作效率，限制端点用户对数据库的使用.关系模型严格符合现代数据模型的定义，其数据结构简单清晰，存取路径完全向用户隐蔽，使程序和数据具有高度的独立性.关系模型的数据语言非过程化程度较高，用户性能好，具有集合处理能力，并有定义、操纵、控制一体化的优点.关系数据库系统为提高程序员的生产率，以及端点用户直接使用数据库提供了一个现实基础.

关系模型(relational model) 数据库中常用的数据模型之一.它把数据组织成二维表的形式,即用单一的二维表的结构表示实体及实体相互之间的联系.二维表应满足以下条件:

1.表中每一项必须是基本数据项.

2.表中每一列是同类型的数据.

3.表中每一列必须具有惟一的名字.

4.表中不应有相同行.

5.行与列的顺序与表中信息无关.

满足上述条件的二维表称为一个"关系".二维表的每一行称为关系的一个元组;每一列称为关系的一个属性.关系中的每一列的值总是取自一个集合,这个集合称为域.对元组集合的结构的描述称为关系模式,它是一个五元组$\langle R,U,D,\mathrm{DOM},F\rangle$,其中:$R$为关系名;$U$为属性集合;$D$为域集合;DOM为属性到域的映射;$F$为一组数据相关,是一类完整性约束条件.某一时刻一个关系模型的实例称为关系状态,简称关系.

关系模型的操作部分具有关系处理能力,它把整个关系作为操作对象.具有关系处理能力的关系数据语言,可分为关系代数、关系演算和介于两者之间的语言.一般以关系代数作为度量语言处理功能的标准.关系代数除提供传统的集合运算外,还提供了选择、投影、连结等操作,从而有利于对复杂关系的数据存取.

网状模型(network model) 数据库系统中最常见的数据模型之一.属于格式化模型(参见"数据模型").网状模型的特征是:

1.可以有一个以上的结点没有亲体(双亲).

2.至少有一个结点有多于一个的亲体(双亲).

网状模型与层次模型不同之处主要有以下两点:

1.层次模型中从子女到亲体的联系是惟一的,而在网状模型中从子女到亲体的联系不是惟一的,因此在网状模型中需将每一种联系(通常称为"系")给予名字.

2.网状模型中允许复合链,即两个记录之间可以有两种或两种以上联系.

层次模型(hierarchical model) 数据库系统中最常用的数据模型之一.属于格式化模型(参见"数据模型").层次模型可用一棵树来表示,其特征是:

1.有且仅有一个结点无亲体,此结点即树的根.

2.其他结点有且仅有一个亲体.

在树中,每一记录只有一个亲体,因此如果记录R_i有子女R_j,则从R_j到R_i的映像是惟一的.所以对于每一个记录(根结点除外)只需指出它的亲体记录,就可以表示出层次模型的整体结构.

范式(normal form) 满足一定基本条件的关系模式.在关系数据库设计中,选用不同的关系模式集合作为数据库模式,其性能优劣大不相同,为了区分数据库模式的优劣,把数据库模式分为各种不同等级的范式.范式可理解成某一种级别的关系模式的集合,人们常称某一关系模式R为第几范式,就表示该关系的某一种级别,一般R为第几范式就写成$R\in x\mathrm{NF}$.例如R是一个关系模式,如果R中每个属性A的值域中每一个值都是不可分解的,则称R属于第一范式,记为$R\in 1\mathrm{NF}$.又如若$R\in 1\mathrm{NF}$,且每一个非主属性完全函数依赖于码,则称R属于第二范式,记为$R\in 2\mathrm{NF}$.

文件(file) 存储数据的基本概念之一.是根据特定目的而收集在一起的有关数据的集合.文件按内部结构可分为记录式文件和流式文件.流式文件内部没有记录的概念,最常见的流式文件是字符流文件,可看成是一字符串.记录式文件是同一类型的记录的集合,可分为定长记录式文件和可变长记录式文件.记录在文件中的排列可能有顺序关系,但除此之外,记录与记录之间不存在其他关系.在这一点上,文件有别于数据库.文件按记录之间的组织方式又可以分成具有逻辑顺序关系的文件和直接文件两大类:

1.具有逻辑顺序关系的文件.其记录之间存在着某种逻辑上的顺序关系,并且认为按此顺序读写各记录是这类文件的主要使用方式.这类文件包括顺序文件、索引文件和索引顺序文件等.

2.直接文件.文件内记录与记录之间没有顺序关系.例如,散列(杂凑)文件.

数据库(database) 一种管理数据的软件系统.是按照一定的数据模型在计算机系统中组织、存储和使用的相关数据的集合.是数据库系统的一个组成部分.数据库数据模型是对数据库中的数据进行的逻辑组织,或用户所看到的数据库的数据间的逻辑结构.数据存储在计算机外部存储器(如磁鼓、磁盘、磁带等)上,数据库具有数据结构化、最低冗余度、较高的程序与数据的独立性、易于扩充、易于编制应用程序等优点.较大的信息系统都是建立在数据库设计之上的.

数据库是计算机软件的一个重要分支,是20世纪60年代后期发展起来的数据管理新技术,最早出现于1968年9月,美国IBM公司发表了"信息管理系统"(IMS)的第一个版本,1970年又出现了"关系数据库".随着20世纪70年代后期分布计算机系统的发展,成功地研究了分布式数据库系统.目前数据库研究的课题,主要涉及数据库管理系统软件的研制、数据库设计、数据库理论等领域.

数据库系统(database system) 计算机的数据管理系统.由数据、硬件、软件和管理人员的总体而

构成的数据管理系统.一个数据库系统除计算机系统的一般的硬件、软件外,还包括以下四个基本组成部分:

1. 直接存取存储设备,用以存储大量的数据.

2. 数据库,按一定数据模型组织的相关数据的集合.

3. 数据库管理系统,负责数据库的建立、使用和维护的软件系统.

4. 数据库管理员,管理数据库的人员.

数据库系统由文件系统发展而来,它具有数据、体系和控制三个方面的主要特征:

1. 数据特征.数据库既描述数据本身,又描述数据之间的联系,是面向数据总体结构的信息集合.它的数据结构反映了某一部门的整体信息结构,数据冗余度小,易于扩充新的应用,可为多个用户共享.

2. 体系特征.数据库模式由存储模式和逻辑模式组成.存储模式描述数据的物理结构;逻辑模式描述数据的逻辑结构,它又分为模式和子模式两种.在数据库系统中,程序和数据之间具有物理和逻辑两级独立性,数据的存储模式改变,而数据库模式可以不改变,因而不必改写应用程序,这称为物理独立性;数据库模式改变时,子模式可能不改变,也就不必改写应用程序,这称为逻辑独立性.这种较高的程序与数据的独立性,有利于应用程序的编制与维护.

3. 控制特征.数据库在定义、建立、运行和维护时是由数据库管理系统来管理和控制的,有利于保证数据库数据的安全性、完整性和并发操作的一致性.

数据库管理系统(database management system) 简称 DBMS.管理并控制数据库的一组软件.它建立在操作系统的基础上,是数据库系统的重要组成部分.用户使用数据库的各种命令以及应用程序的执行,都要通过 DBMS. DBMS 的功能随系统而异,一般地,包括以下几个方面:

1. 数据库定义,包括逻辑数据结构定义、存储结构定义、保密定义以及信息格式定义等.

2. 数据库管理,包括系统控制、数据存取及更新管理、数据完整性及安全性控制、并发控制等.

3. 数据库建立和维护,包括数据库的建立、更新、再组织、恢复、数据库结构维护以及性能监视等.

4. 通信,具备与操作系统的联机处理、分时系统及远程作业输入的相应接口.

DBMS 通常由数据库语言和数据库管理例行程序等组成:

1. 数据库语言,包括数据描述语言及其翻译程序和数据操纵(或查询)语言及其编译(或解释)程序.

2. 数据库管理例行程序,一般包括三部分:系统运行控制程序、语言翻译例行程序以及 DBMS 的公用程序,由它们完成 DBMS 的功能.

目前商品化的数据库管理系统已有许多,如 dBASE, FoxBASE, Ingres, Informix, Oracle DB2, SQL 等.

面向对象数据库系统(object-oriented database system) 一种新的数据库管理系统.随着计算机应用领域的扩展与深化,要求数据管理软件管理复杂对象,并模拟复杂对象的复杂行为.于是,在 20 世纪 80 年代后期产生了面向对象数据库系统,把面向对象技术与数据库技术结合起来,利用类的设施来描述复杂对象,利用类中封装的方法来模拟对象的复杂行为,利用继承性来实现对象的结构和方法的重用.在面向对象技术与数据库技术相结合的过程中,基本上是沿着两种途径发展的:一种是以某一种面向对象语言为基础,把数据库的某些功能加到面向对象语言中去,使其成为一个面向对象数据库管理系统(即 OODBMS).另一种途径是将传统的关系数据库加以扩充,增加面向对象特性,建立对象——关系数据库管理系统(即 ORDBMS).前者主要支持持久对象和实现数据共享,但不支持关系数据库标准语言 SQL,在通用性与应用领域方面受到很大的局限,后者很好地解决了这些不足.

数据库设计(database design) 设计数据库的结构和建立数据库的过程.即在计算机系统具有的数据库管理系统的基础上,按照应用要求以及计算机系统所提供的数据模型和功能,设计一个结构良好、使用方便、效率较高的以数据库为核心的应用信息系统过程.数据库设计既要考虑多种用户对共享数据的不同要求,同时又必须综合平衡影响系统性能的各种工程因素.

数据库设计过程可分为环境调查与系统分析、建立概念数据模型、逻辑模式设计和物理模式设计等四个基本阶段.数据库设计方面的主要研究方向是为数据库设计提供能够有效地表达语义关系的数据模型;为设计各阶段提供交互式的设计工具,以及把设计工具一体化,以构成完整的数据库设计研制环境等.

模式(schema) 数据库数据的全局逻辑结构.数据库数据有物理的和逻辑的两个侧面.数据物理结构的描述称为存储模式.数据逻辑结构的描述称为逻辑模式.逻辑模式又分为全局模式(简称模式)和子模式.前者描述全局逻辑结构,后者描述某些应用所涉及的局部逻辑结构.存储模式和逻辑模式总称为数据库模式.模式是数据库所有数据元素类型的一个结构图,用模式数据描述语言来定义.它是数据库的一种描述而不是数据库数据本身,仅是装入数据的一个框架.

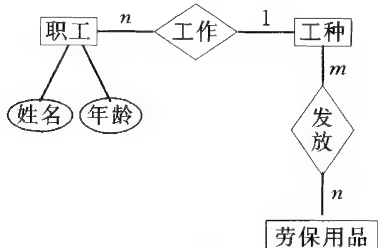
子模式(subschema)　数据库数据的局部逻辑结构.数据库数据具有物理的和逻辑的两个侧面,数据库中描述数据物理结构的为存储模式,描述数据逻辑结构的为逻辑模式,逻辑模式又分为全局模式(参见“模式”)和子模式.子模式是相应模式的子集,但可以在一定范围内有所变化.由一个模式可以构成许多个子模式,子模式之间可以互相覆盖.子模式定义了与某用户的一个或多个应用程序有关的数据元素的名字、特征及其相互关系,是用子模式数据描述语言描述的.同一子模式可以供任意多个应用程序共享,但一个应用程序只能启用一个子模式.子模式的概念是很重要的.首先,程序员不必关心整个数据库而只关心数据库中与其有关的那一部分数据,从而简化了程序的编写、调试和维护;其次,程序被局限于子模式,且只有通过子模式才能使用模式,从而自动保证了数据库中与子模式无关的其余部分的安全性和完整性;此外,子模式的概念提高了程序和数据的逻辑独立性.

实体联系图(entity-relationship diagram)　简称E-R图.表示信息结构的框图.在数据库逻辑设计的过程中,用它来简明地表示现实信息的结构实体联系图,是陈(Chen,P. P. S.)于1976年提出的.在E-R图中使用了以下的一些记号(如图):

1.用矩形框表示实体,在框内标明实体名.

2.用菱形框表示实体间的关系,在菱形框内写上关系名,并用无向直线将菱形框分别与有关的实体相连结.

3.用椭圆形表示实体的有关属性,并用无向直线与它所描述的实体相连结.



E-R 图

操作系统(operating system)　系统软件的重要组成部分.负责管理和控制计算机系统的各种软硬件资源,以便合理、高效、方便地利用计算机系统.操作系统是用户和计算机之间的接口,用户通过操作系统使用计算机.操作系统主要提供以下功能:

1.处理器管理,对处理器的使用进行调度.

2.存储管理,分配和回收存储单元.

3.设备管理,控制外围设备操作,分配和回收外围设备资源.

4.作业管理,调度和控制作业的执行.

5.信息管理,提供存取信息的各种服务.

操作系统可分为六种类型:

1.单用户操作系统.

2.批处理操作系统.

3.分时操作系统.

4.实时操作系统.

5.网络操作系统.

6.分布式操作系统.

存储管理(memory management)　由操作系统提供的一种功能.它能分配和回收存储单元.一道程序在启动执行前必须装入存储器内;程序在执行中可能还要申请存储单元;程序执行完毕后要将所占用的全部存储单元归还给系统.这些存储单元的分配和回收工作,都由操作系统来完成.具有多道程序设计功能的操作系统,还必须采取有效的策略,分配各用户程序所需的存储资源,以提高系统效率.在采用页式存储或虚似存储的系统中,操作系统还要配合硬件做页面调度工作,即将所需的页面调入存储器,将暂时不用的页面调出以腾出空间.页面调动工作完成的好坏,对系统效率有很大的影响.

处理器管理(processor management)　操作系统提供的一种功能.它能对处理器的使用进行调度.在具有多道程序设计功能的操作系统中,处理器在操作系统控制下交替地为各道程序所占用.操作系统根据一定的调度算法分配处理器.常用的处理器调度算法有优先数法和时间片法等.

1.优先数法.每一道程序给定一个优先数,每次分配处理器时,优先数大者优先分配,只有当优先数较大的那些程序因为等待外围设备传输或其他原因暂不能占用处理器运行时,优先数较小的程序才有可能占用处理器.批处理操作系统常采用这种调度算法.

2.时间片法.每道程序分得一个时间片,当一道程序连续占用处理器的时间超过一个时间片时,就强行把处理器分配给另一道程序.分时操作系统常采用这种调度算法,以保证各终端的用户有均等的占用处理器的机会.

设备管理(device management)　操作系统提供的一种功能.它控制外围设备操作,分配和回收外围设备资源.当一道程序在执行中需要使用外围设备时,它向操作系统提出使用要求.操作系统根据要求,启动外围设备并控制其操作.从分配的角度看,外围设备可分为共享设备和独占设备两类:

1.共享设备.指可以为多个程序所共享的设备,如磁盘等.对于共享设备的分配,必须有一定的策略来决定谁可占用,占用多少.

2.独占设备.指一次只能分配给一道程序的设备,如行式打印机等.对于独占设备,可以按一定的

策略分配，也可以设置虚拟设备而不分配. 例如，将行式打印机设置成虚拟设备后，打印输出操作就由操作系统转换成写入磁盘操作，即将要打印的信息存入磁盘内，到程序执行完毕或适当的时候，才由操作系统控制而从行式打印机上输出. 这就是所谓的假脱机输入输出或斯普林方法. 采用这种方法可以增大系统的吞吐量，提高系统的效率.

文件管理(file management)　由操作系统提供的一种功能. 它具有存取信息的各种服务的功能. 它主要由操作系统中的文件管理系统来完成. 提供的文件服务，主要是从一个指定文件读出信息、修改指定文件、建立和废除一个文件等. 用户通过文件系统使用文件时，不需要考虑信息存放的物理位置，也不需要了解信息如何存储在物理介质上，而只需知道文件的标识名字和信息的逻辑结构. 此外，用户还可以通过文件管理系统来规定文件的使用权限，如规定哪些用户拥有阅读某文件的权利、谁还拥有修改的权利、谁拥有废除的权利等. 因此，在文件管理系统管理下，各种文件不仅使用方便，而且安全可靠.

作业管理(job management)　操作系统中的一种功能. 它具有调度和控制作业的执行的功能. 操作系统控制作业的执行的方式有两种：一种是在用户输入作业之前先用系统提供的作业控制语言(JCL)编写作业说明书，指明系统如何控制作业的执行，然后将作业说明书输入计算机. 例如，作业说明书可指出，将指定的源程序用某种编译程序编译并随后执行. 另一种方式是用户从终端打入各种命令，告诉操作系统如何控制作业的执行. 操作系统也把执行中的情况通过终端向用户报告. 计算机系统常常同时接受多个作业，操作系统必须根据一定的策略来调度作业的执行. 当采用多道程序设计方法来设计一些较大型的操作系统时，在系统中要经常保留多个运行的作业. 从系统已接纳的一批作业中挑出若干个可运行的作业，并为这些被选中的作业分配所需的系统资源，以达到较佳的系统效率，更好地为用户服务，这个过程称为作业调度.

作业(job)　操作系统的基本概念之一. 是用户要求计算机系统处理的任务. 例如，要求计算机系统执行一个或多个程序就是一个作业. 作业是计算机工作任务的单位，一个作业可能包括几个程序的相继执行(例如，以高级语言写成的程序的编辑、编译及执行)；也可能需要同时执行为同一任务而协同工作的若干程序(例如，其中一个程序控制打印数据，另一个则计算并产生输出). 作业管理是由操作系统来完成的.

计算机系统常常同时接受多个作业，操作系统必须根据一定的策略来调度作业的执行. 在操作系统控制下，作业一般分为成批作业和分时作业两类. 前者包括后台作业、脱机作业和非会话型作业；后者包括前台作业、联机作业和会话型作业等.

进程(process)　操作系统的基本概念之一. 进程有两个含义：

1. 进程是操作系统的基本概念之一. 它相当于计算机中某个程序的一次执行. 程序指的是计算机为了获得某种结果而遵循的一个指令序列，是一个静态的概念. 进程则指程序的执行过程，是一个动态的概念. 用同一程序对不同的数据先后或同时加以处理，就相当于有好几个进程.

2. 进程或称路径，或称任务，是并发程序出现后出现的一个重要概念. 在并发程序设计中，可同时执行的程序模块称为进程(参见“并发程序设计”).

进程有两个侧面：

1. 进程说明，用以定义一个行为模式，一般包含一组数据和一组操作(顺序操作).

2. 进程语句，用以按照相应进程说明中所定义的模式来启动一组动作的执行.

死锁(deadlock)　操作系统的基本概念之一. 指进程因争夺资源而无休止地相互等待. 例如，进程 P_1 占有了绘图机而申请行式打印机，进程 P_2 占有了行式打印机而申请绘图机. 它们都因为申请不到资源而永远等待，这就是死锁. 在死锁的状态下，计算机系统效率将大大下降，因此，死锁的情况应尽量加以避免. 解决死锁问题有两种途径：一是预防死锁，在操作系统中设计各种资源调度算法，防止死锁发生；二是检测死锁，当死锁发生时能及时发现并进行排除.

分时处理(time-sharing processing)　一种作业处理方式. 是多个用户可同时与计算机系统交互处理作业的方式. 分时处理是将主计算机 CPU 的运行时间分割成一个个长短相等(或基本相等)的微小时间片，把这些时间片依次轮流地分配给各个终端用户程序，每个用户程序仅仅在其获得该 CPU 时间片内执行. 当时间片完结，用户程序又处于等待状态，此时，CPU 又在为另一个用户服务. 用户程序就这样断断续续、停停走走，直到最终完成执行. 虽然在微观上(微小时间片的数量级)用户程序的执行是断续的，作业运行是不连续的，但在宏观上，用户的任何请求服务总能够及时得到响应.

分时处理具有多路调制性、交互性、独占性三个特点. 所谓多路调制性是指同时允许多个分时用户工作；交互性是指用户在终端上与主计算机交互会话来联机控制自己的作业运行；独占性是指用户之间彼此互不干扰，每个用户都感觉好像自己在单独使用计算机.

磁盘操作系统(disk operating system)　亦称 DOS 操作系统. 一种操作系统. 是用于微型计算机

的操作系统. DOS 操作系统是运行在个人计算机上的操作系统,自 1981 年第一版问世以来,已发展到第七版,其间所发行的各种改进版、更新版、增强版和不同地区及语言的版本多得不计其数. 在应用领域中形成了两个完全相兼容的 DOS 版本,即 MS-DOS 和 PC-DOS,前者是微软(Microsoft)公司的产品,后者是 IBM 公司的产品. DOS 操作系统采用层次化结构,其系统核心位于三个系统文件上,它们是:系统初始化,即 I/O 基础程序;内核初始化,即系统调用程序;命令处理程序,每一部分对应用程序提供特定的支持. DOS 操作系统随着硬件的发展,已经从简单的单用户、单任务操作系统,发展到能支持高级文件操作、支持局域网操作、支持简单的多任务切换和具有一定的彩色图形用户界面的高级单机操作系统. DOS 操作系统自第七版以后就停止了再开发,逐渐地让位于 Windows 操作系统.

UNIX **操作系统**(UNIX operation system) 一种操作系统. 是一个通用的、多用户的计算机分时操作系统. UNIX 操作系统是美国贝尔实验室的汤普森(Thompson, K.)和里奇(Ritchie, D. M.)于 1969 至 1970 年研制的,至 1982 年已产生了第七个版本,现已成为能在各种不同类型的计算机上运行的标准操作系统. UNIX 操作系统功能简单实用,结构灵活,使用方便. 它的成功主要是采用了一系列先进技术和措施,解决了一系列软件工程的问题. 它提供了许多在大型操作系统中很少看到的特性. 例如:

1. 有一个能编入可移动卷的分级结构文件系统.

2. 文件、设备和内进程输入输出可兼容.

3. 具有启动同步进程的能力.

4. 每个用户可按自己的愿望选择系统命令语言.

5. 高度可移植性.

6. 提供十几种程序设计语言和其他大量(现已超过 200 个)的实用程序,UNIX 操作系统较新的两个版本是 1980 年推出的系统 Ⅲ,它加进了网络功能,综合了 UNIX 和 PWS(程序员工作台)操作,1983 年推出的系统 Ⅴ,着重于提高操作效率.

UNIX 操作系统由核心与外层两部分组成. 核心部分包括了操作系统的主要功能,如存储 Shell 管理、进程与处理机管理、设备管理与文件管理等. 外层由 Shell 与系统调用两部分构成,Shell 被称为 UNIX 的外壳,是用户与系统交互的环境与接口,系统调用是用户程序与系统核心的接口.

Windows **操作系统**(Windows operation system) 亦称视窗(Windows)操作系统. 一种操作系统. 是微软公司推出的一个采用图形用户界面(GUI)的新颖的操作系统. Windows 最初是作为对 DOS 的图形化扩充而推出的. 第一个图形用户界面系统是苹果公司 1984 年推出的 Apple Macintosh. 自微软公司 1985 年推出 Windows V1.0 版以来,功能得到了极大的改进,尤其是 1990 年推出的 WindowsV 3.0 版奠定了视窗操作系统的基础,直至 1995 年推出的视窗 95(WIN 95)更是确立了视窗操作系统在个人计算机上的主导地位. 目前已有 WIN 98, WIN 2000 的视窗操作系统版本. Windows 操作系统由应用界面与系统核心两部分组成,应用界面由程序管理器、任务管理器、文件管理器与控制面板四部分组成,Windows 操作系统核心由用户接口服务、系统服务、图形设备服务三个模块以及独立的文件系统组成. Windows 操作系统提供了保护模式、多任务、动态连结、窗口管理、图形界面等多种系统服务,提供了一套与设备独立的I/O驱动程序接口和图形函数,形成良好的用户环境.

Linux **操作系统**(Linux operating system) 一种新型的网络操作系统. 1991 年 8 月,芬兰赫尔辛基大学生林纳斯(Linus, B. T.)在因特网上发布了他为 386 或 486 开发的 Unix-like 操作系统,免费让网友使用. 由于它是 Linus 的 Unix,故名 Linux. 顾名思义,它有许多类似 Unix 的功能(参见"Unix 操作系统"). 然而,它有两个明显的特点:一是公开源代码,允许人们利用源代码进行开发、改进缺点、增加性能,因此,它的可靠性与安全性得以提高,比封闭源代码的 Windows 和 NetWare 高得多. 二是免费下载,降低了成本,从而占领了客户机/服务器的不少市场. Linux 的出现,为打破软件垄断提供了崭新的途径.

多媒体技术(multimedia technique) 一种计算机技术. 是利用计算机来交互地综合处理文字、图形、图象、动画、音频和视频等多种媒体信息,并使这些信息建立逻辑连结的一种计算机技术. 所谓"媒体"是指消息的载体,包括文字、图形、图象、动画、音频和视频等,多媒体就是综合集成两种以上媒体,共同表示、传播和存储同一信息的新媒体. 多媒体技术的应用领域十分广泛,多媒体视频点播系统、多媒体视频会议系统、多媒体监视系统、多媒体远程教学系统、多媒体远程医疗系统、多媒体数据库系统、电子商务系统、电子出版物以及多媒体家电等都是多媒体技术的具体应用.

多媒体软件(multimedia software) 一类计算机软件. 多媒体技术涉及种类繁多的各种硬件,要处理大量的多媒体数据,就要将多媒体技术涉及的各种硬件有机地组织起来,进行有效地管理;使用户方便有效地控制、管理和使用各种多媒体数据,这些软件称为多媒体软件. 多媒体软件有:

1. 多媒体核心软件. 包括多媒体设备驱动程序

和多媒体操作系统.

2. 多媒体工具软件. 包括多媒体数据处理软件和多媒体编辑创作软件.

3. 多媒体应用软件. 包括面向诸多应用领域的各种实用性软件系统.

软件工程(software engineering) 用工程方法研制和维护软件的过程和有关技术. 软件研制包括需求定义、设计、实现和测试四个阶段;软件维护指使用软件的过程中对软件进行的修改和完善.

软件工程是在 20 世纪 60 年代后期人们开始认真研究解决软件危机的方法而逐步形成的计算机科学技术领域中的一门新兴学科. 20 世纪 60 年代中期,软件的复杂程度迅速提高,研制周期很长,正确性难以保证,可靠性问题非常突出,即出现了所谓"软件危机",其实质是人们难以控制软件的研制和维护. 面临这种形势,人们意识到靠传统的方法再也无法驾驭极为复杂的软件生产,于是有人提出了用系统工程学的原理和方法来管理软件生产过程的设想. 到 1968 年,在北大西洋公约组织(NATO)主办的会议上才正式提出了"软件工程"的术语.

软件工程的主要对象是大型软件. 软件工程研究的问题主要有:质量保证和质量评价;研制和维护的方法;工具系统、文件、用户界面的设计;软件管理等. 软件工程的最终目的是摆脱手工生产软件的状况,实现软件研制和维护的自动化.

软件生存周期(software life cycle) 软件工程中的一个重要概念. 软件产品从形成概念开始,经过研制、交付使用,在使用中不断增补修订,直到最后被淘汰,让位于新的软件产品的过程称为软件生存周期. 一个软件产品的生存周期可划分成若干个相对独立而又彼此联系的阶段. 每一阶段的开始和结束都有严格标准,对于任何两个相邻的阶段而言,前一阶段的结束标准就是后一阶段的开始标准. 每一阶段中的工作,均以前一阶段任务的完成为依据,并为后一阶段的工作创造前提和基础. 每个阶段都采用科学的管理技术和良好的技术方法,在每个阶段结束之前都从技术和管理两个角度进行严格的审查,合格之后才开始下一阶段的工作. 这就使软件工程的全过程以一种有条不紊的方式进行,从而有利于保证软件质量,降低软件成本,提高软件开发的成功率和生产率.

软件生存周期的阶段划分尚无统一的标准. 一般地,软件生存周期由软件定义、软件开发和软件维护三个时期组成,每个时期又进一步划分成若干个阶段:

1. 软件定义时期. 主要任务是确定软件产品的总目标、功能、可行性、成本和实施计划,通常划分成问题定义、可行性研究和需求分析等阶段.

2. 软件开发时期. 任务是具体设计和实现在前一时期定义的软件,它通常包括一般设计、详细设计、编码和测试等阶段.

3. 软件维护时期. 主要任务是使软件持久地满足用户的需要. 通常不再划分阶段.

结构化分析(structured analysis) 简称 SA 方法. 一种简单实用的系统分析方法. 是由美国 Yourdon 公司提出的. 其基本思想是"自顶向下逐层分解". 使用这种方法,无论系统多么复杂,分析工作都可以有计划、有步骤、有条不紊地进行,系统规模再大,分析工作的复杂程度不会随之增大,而只是多分解几层而已,所以 SA 方法有效地控制了复杂性. SA 方法是软件需求分析阶段使用很广的一种分析技术,它适用于分析大型数据处理系统,特别是企事业管理方面的系统. SA 方法通常与设计阶段的结构化设计(SD)方法衔接起来使用. 目前大多数计算机系统(称为目标系统)都是用来代替一个当前已经存在的人工数据处理系统(称为当前系统),对于这类系统的分析过程,SA 方法一般按以下四步进行:

1. 理解当前的现实环境,获得当前人工系统的具体模型.

2. 从当前系统的具体模型抽象出当前系统的逻辑模型.

3. 分析目标系统与当前系统逻辑上的差别,建立目标系统的逻辑模型.

4. 对目标系统的逻辑模型进行补充、完善.

使用 SA 方法获得的系统规格说明书由以下几个主要部分组成:

1. 一套分层的数据流图,描述系统的分解,即描述系统由哪些部分组成,各部分之间有什么联系等.

2. 一本数据词典,定义数据流图中的所有数据流、文卷和数据项等数据特征.

3. 一组加工说明,精确描述数据流图中的基本加工(处理),加工逻辑一般用结构化语言、判定树、判定表等形式描述.

4. 其他补充材料等.

结构化设计(structured design) 简称 SD 方法. 一种系统设计方法. 这种方法适用于任何软件系统的总体设计,可以与软件工程中分析阶段的结构化分析(SA)方法以及编程阶段的结构化编程(SP)方法前后衔接起来使用. 其主要特点是:在 SA 方法的基础上,根据系统规格说明书的数据流图,采用模块化的方法,自顶向下地将系统设计成由相对独立、单一功能模块组成的程序结构,从而由逻辑模型得出具体的设计方案.

结构化设计的主要工具是控制结构图. 它自顶向下地反映出软件系统各部分的组成、相互关系及相互影响. 由它即可得出数据存储的安排及软件编

制计划等. SD 方法的优点是易于避免接口错误和所谓“波及影响”(修改一处,影响全局),从而使系统结构清晰,不但易于分工编写,而且提高了系统的质量.

面向对象分析(object-oriented analysis) 一种新的系统分析方法. 它是采用面向对象的风格进行系统分析和需求定义的重要方法. 面向对象分析的关键在于理解问题空间并将其模型化,它采用人们认识客观事物和理解现实世界过程中常用的基本法则:认识对象及其属性;认识对象的整体及其组成部分;对象的形成及类的区分;对问题空间进行理解并抽象成模型. 面向对象分析有五个阶段:标识对象、标识结构、标识主题、定义属性、定义服务,即分五个层次建立面向对象分析的模型. 面向对象分析的优点是使功能分析与数据分析使用统一的概念和方法,克服了结构化分析中两者之间的不一致性和不协调性.

面向对象设计(object-oriented design) 一种新的系统设计方法. 面向对象设计是在面向对象分析的基础上进行的,它针对与系统实现有关的因素,利用面向对象分析得到的五个层次构成的模型进行设计,针对实现的要求进行必要的增补和调整. 面向对象设计有四个部分:问题部分、人机交互部分、任务管理部分、数据管理部分. 面向对象设计模型从横向看是上述四个部分,从纵向看每个部分仍然是五个层次. 面向对象设计的优点是使分析与设计有机地相结合,避免了很多重复性的工作,明显地减少了工作强度和文档数量.

软件测试(software testing) 为发现软件中的错误而执行程序的过程. 软件测试只能用于查找程序中的错误,不能证明程序中没有错误. 由于测试的目标是暴露程序中的错误,从心理学角度看,由程序的编写者自己进行全部测试是不恰当的,在综合测试阶段通常由其他人员组成的测试小组来完成测试工作. 软件测试是保证软件质量的关键,它是对需求分析、设计和编码的最后复审. 因此,软件工程要重视测试.

软件测试的方法一般有黑盒测试和白盒测试两种. 对于前者已经知道了软件产品应该具有的功能,可通过测试来检验是否每个功能都能正常使用;对于后者,知道软件产品内部工作过程,可以通过测试来检验产品内部动作是否按照规格说明书的规定正常进行. 软件测试过程与开发过程类似,必须分步骤进行,每个步骤在逻辑上是前一步骤(若有的话)的继续,也是后一步骤(若有的话)的前提. 大型软件系统的测试基本上由模块测试、集成测试和验收测试三个步骤组成.

软件维护(software maintenance) 修改软件的过程. 在软件已经交付使用之后,为了改正错误或满足新的需要而修改软件的过程. 软件维护同一般意义下的维护具有本质的区别. 软件维护有如下四方面的含义:

1. 校正性维护. 诊断和改正软件在研制过程中所潜藏的错误.

2. 适应性维护. 为适应系统运行环境条件的改变而对软件进行的改动.

3. 完善性维护. 对软件的功能加以扩充,使之进一步完善.

4. 预防性维护. 为了改进软件未来的易维护性和可靠性,或为了给未来的改进提供一个更好的基础而对软件进行的修改. 在一个实际的系统中,维护总是不可避免的,且维护费用占软件总预算的绝大部分. 因此,在软件的研制阶段就必须考虑到软件的易维护性,以减少维护的工作量.

软件开发工具(software development tool) 一种计算机软件工具. 是辅助和支持其他软件的定义、开发和维护的工具,简称软件工具. 使用合适的软件工具可以显著提高软件生产率,改善软件质量. 软件生存周期中的各个阶段使用的软件工具有:

1. 需求定义工具. 需求定义阶段所使用的工具,包括描述工具和分析工具.

2. 设计工具. 设计阶段所使用的工具,包括描述工具和变换工具. 描述工具有图示工具(如流程图、框图等)、表格工具(如抉择表)和语言工具(如 PDL、伪码等).

3. 实现工具. 实现阶段使用的工具,主要包括各种程序设计语言、编辑程序、翻译程序、实用程序、模拟程序、仿真程序等.

4. 测试和验证工具. 测试阶段所使用的工具,如静态分析工具、覆盖监视工具、测试数据产生程序、跟踪程序、排错程序、测试程序和验证程序等.

5. 管理工具. 提供给管理人员使用的自动管理工具,如研制进度表、评价统计等.

软件工具的发展趋势是品种多样化、接口标准化、描述形式化、功能智能化和操作自动化.

人 工 智 能

人工智能(artificial intelligence) 计算机科学的一个分支. 是研究、解释和用计算机来模拟人类智能、智能行为及其规律的一门学科. 自从计算机问世以来,人们就开始了人工智能的研究. 从 20 世纪 50 年代到 60 年代初,已编写出一些试验性程序,建立了专用语言——表处理语言. 20 世纪 60 年代末以来,又出现一些具有实际使用价值的系统. 20 世纪 70 年代陆续研制出运用专门知识,特别是经验知识

来解决专门领域难题的专家系统. 1977 年提出的知识工程促进了人工智能从单纯的理论研究走向实用化. 20 世纪 80 年代出现了研究人工神经网络的热潮. 20 世纪 90 年代引入了其他学科的思想和研究方法,如基于自然选择的遗传进化模型、计算智能等. 人工智能研究的主要内容包括知识工程、自然语言处理、机器学习、智能机器人、模式识别、神经计算、自动程序设计、智能控制、智能决策支持系统等.

1. 知识工程. 研究基于知识的计算系统的整个开发过程及有关技术,包括知识获取、知识表示、自动推理和启发式搜索等内容.

2. 自然语言处理. 用计算机处理、加工人类的书面和口头形式的自然语言的过程和有关技术.

3. 机器学习. 为计算机提供自学习能力,使系统可以接受教育、获取知识,学会完成工作的方法和提高工作效率.

4. 智能机器人. 具有人类所特有的某种智能行为的机器.

5. 模式识别. 利用计算机或其他装置,对物体、图象、声音、文字等模式信息进行自动鉴别、分类和解释的过程和有关技术.

6. 人工神经网络. 由大量神经元互连组成的非线性、自适应信息处理系统.

知识工程(knowledge engineering) 知识信息处理的过程及有关技术. "知识工程"这个术语是由美国斯坦福大学的费根鲍姆(Feigenbaum, E. A.)在 1977 年于麻省理工学院召开的人工智能国际会议上提出的. 它是从人工智能派生出来的一个研究领域. 知识工程的研究促进了人工智能从单纯的理论研究走向实用化,并向计算机科学以及其他学科的许多研究领域中渗透,所以有时又把知识工程称为应用人工智能.

知识工程把有关知识库系统、专家系统等的构造技术作为主要课题. 其研究的主要领域包括以下三个方面:

1. 知识获取. 研究知识处理系统如何从系统外部获得知识、充实知识库,包括对外部的数据进行知识化.

2. 知识表示. 研究怎样对知识进行形式化地描述,以便让计算机能合理地存贮和使用知识.

3. 知识使用. 研究在知识处理系统中应如何组织和利用知识,使用怎样的推理方法,以达到所希望的目标.

知识表示(knowledge representation) 知识的形式化表达. 它能使计算机利用知识进行智能信息处理. 研究知识从自然记载形式转变成适合于计算机处理的表示形式. 知识表示一般应具有下列特性:

1. 广泛的知识表示能力,具有直接描述问题世界的表示能力,可表示具有结构的知识、不完全的知识以及元知识等.

2. 模块性,便于知识的追加、修正及阶段性扩展.

3. 易于管理,易检测出矛盾的知识、冗余的知识等,以维护知识库的一致性.

4. 易于人们理解.

5. 与高效的推理方法相配合,例如应具有联想功能等.

知识的表示方法大体上可分为过程型表示法和陈述型表示法两大类. 过程型表示法是根据要解决的特定问题,指出一系列操作过程的表示方法,其执行效率高,但适应性差. 陈述型表示法是将事实和判断规则分别加以说明的表示方法,它可用知识表示相互独立的事物,易于追加、修正,具有较好的通用性和适应性,但处理问题的效率较低. 目前使用的具有代表性的知识表示方法大致有以下几种:

1. 产生式系统,用产生式规则来表示知识的专家系统.

2. 黑板模型,不同层次、不同类的知识源,通过被称为黑板的公用数据区而得到协调的处理.

3. 框架系统,按照框架理论建立的知识表示模型.

4. 谓词逻辑,用逻辑式的集合来记述知识,并含有谓词逻辑的推理功能,如基于归结原理的演绎推理机能.

5. 语义网络,由结点和连结结点的有向弧构成,具有对网络进行搜索和操作的算法.

6. 对象模型,是表达以解决知识库的问题为对象的系统及概念的模型,能某种程度地使用深层知识.

7. 上述几种方法的某种组合.

产生式系统(production system) 一种专家系统. 是用产生式规则来表示知识的专家系统. 产生式系统通常由规则库、推理程序和综合数据库三个主要部分组成:

1. 规则库. 管理所有产生式规则. 产生式规则的形式是"如果:〈前提〉,则:〈后果〉".

2. 推理程序. 其任务是运用控制策略找到可以应用的规则. 采用的策略有正向推理(或称"数据驱动型推理"或"自底向上推理",它是从综合数据库初始状态所适用的规则开始,探求达到目标状态的道路)、逆向推理(亦称"面向目标的推理",或称"自顶向下推理",它是从目标状态出发,选择所需要的规则,组成假设或子目标,这样重复操作,看最终的假设是否满足现存的综合数据库的状态)以及双向推理(正向推理和逆向推理并用的一种混合型方法)等.

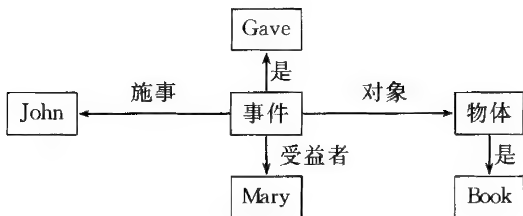
3. 综合数据库. 包含有关求解问题的初始事实和断言，以及推理的中间结果. 当数据库中的事实与某条规则〈前提〉相符合时，这条规则被触发，其〈后果〉中所指示的行动或事实存入综合数据库.

产生式系统的优点是知识的模块性好、可读性好、结构简单，所以被广泛地用于专家系统中. 它的缺点是知识间的关系和相互作用不好表达，且规则越多，规则库的管理越困难，推理效率越低，所以不能面向大规模应用.

框架(frame)　一种表达知识的机制. 是由若干个结点和关系构成的网络，用来表示某类情景的结构化的一种数据结构. 相关的框架可以组合成框架系统. 框架理论是美国麻省理工学院明斯基(Minsky,M.)在 1975 年初提出的，是以与人们的记忆以及情景理解和谈话理解有关的认知科学概念来考察问题的理论. 人们的记忆以框架结构为单位进行结构化，当某种情景出现时，人们就以记忆中的从某种与其相应的场面得到的信息为线索，选出适当的框架，对详细部分的一致性进行推理. 如果不完全一致时，就按这个框架所具有的信息来选择其他合适的框架，反复进行这样的匹配，即利用框架所具有的程序知识来进行这些推理处理. 把按这种理论而建立的知识表示模型称为框架模型. 在框架系统中，知识的表示单位是框架. 框架是以物体或状态等为对象用一定结构来表示的数据结构，由称为“槽”的构成元素的集合所组成，而且在槽中填入与此框架所表示的对象有关的具体的信息值. 框架系统中的推理是根据框架间的消息交换来实现的. 框架系统的特征是：知识表示单位是比较大的集合；使用分类学的层次结构、陈述型知识和过程型知识相配合的表示；使用没有特定的推理控制机构的、被称为框架之间的信息通信的目标指向方法来进行推理等. 这种知识的表示及推理控制的作法具有较大的灵活性，因而多被作为人工智能或者知识工程研究用的软件环境来利用.

框架模型(frame model)　见“框架”.

语义网络(semantic network)　一种用图来表示知识的结构化方式. 在一个语义网络中，信息被表达为一组结点，结点通过一组带标记的有向直线彼此相连，用于表示结点间的关系. 在人工智能的程序中，谓词及其变元可以看做是语义网络中的结点；而格关系则相当于结点之间的连结形式. 语义网络是一种面向语义的结构，它们一般使用一组推理规则，规则是为了正确处理出现在网络中的特种弧而专门设计的. 语义网络对表达典型的陈述句子的内容尤其有用. 例如，语句 John gave the book to Mary 的语义网络表示如下图.



语义网络的一个重要特性是属性继承. 凡用有向弧连结起来的两个结点有上位与下位关系. 例如“兽”是“动物”的下位概念，又是“虎”的上位概念. 所谓“属性继承”指的是凡上位概念具有的属性均可由下位概念继承. 在属性继承的基础上可以方便地进行推理是语义网络的优点之一.

知识获取(knowledge acquisition)　获取信息的方法. 广义地讲，就是收集信息的功能，其中包括以心理学为中心，很早就开始研究的学习功能. 狭义地讲，是指知识信息处理系统如何从外部获得知识，它是知识工程中一个重要的分支，也是最困难、最不成熟的一个分支. 一个困难的问题是如何将人类智能机械地算法化，如何从理解入手，并通过推理(演绎推理和归纳推理)使其形式化以便使用.

知识获取大致可分为人工直接获取和计算机自动获取两类方法.

1. 人工直接获取知识. 由知识工程师沟通领域专家和计算机系统后，抽取、组织知识并实现编码. 这是现在大多数知识工程中所采用的方法. 这种方法首先规定知识表示形式，然后将领域专家的知识变换成这种形式并存入计算机. 全部过程依靠人的手工进行，效率较低.

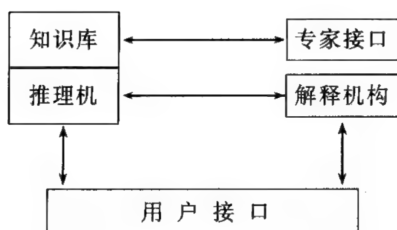
2. 计算机自动获取知识，又称机器学习. 指计算机通过记忆、告知、实例、类比、观察和发现等方式，对外部的数据进行演绎推理和归纳推理，以获取、组织和扩充知识，从而改善系统自身的性能. 由于获取知识使用的归纳推理要比演绎推理更困难，所以具备自动学习功能的知识获取工具目前还处于研制阶段.

专家系统(expert system)　亦称基于知识的系统. 计算机的一种程序系统. 它是能像某一领域专家那样向用户提供解决问题的办法的计算机程序系统. 专家系统把有关领域的专家知识表示成计算机能够利用的形式，通过推理使计算机能够以和专家相近的水平，解决该领域原来只有专家才能解决的复杂问题. 专家系统的基本组成如图所示. 它包括有：

1. 知识库，存放按要求表示的专家知识，并对其进行管理.

2. 推理机，利用知识库内的知识进行推理以得出结论.

3. 用户接口，同用户进行友好对话.



专家系统基本组成

4. 专家接口，从专家那里获得专业知识或经验知识以建立知识库.

5. 解释机构，说明所给结论的根据，以及在此结论下采取哪些对策，便于用户理解结论含义.

其中，以知识库和推理机最为重要，是专家系统的核心. 专家系统除了使用产生式规则以外，有的也使用语义网络、框架等表达知识，虽然它们更复杂些，但能更好地表示许多事物之间的联系. 专家系统的最大特点在于知识库同推理机构是相互独立的，使知识库可以通过教学而不断更新和补充新知识，不必修改程序.

基于知识的系统(knowledge-based system) 即"专家系统".

专家系统开发工具(meta expert system) 一种计算机软件. 它可以为各个领域的专家提供开发专家系统的环境和工具，使得他们在研制专家系统时尽可能少地依赖知识工程师，而通过"元专家系统"直接开发建造"专家系统"的软件程序. 与通过知识工程师用人工智能语言直接编写程序相比，它可以缩短专家系统的开发周期，提高专家系统的启发性、透明性、结构性，改善专家系统的人-机界面，提高知识获取过程的自动化程度.

自然语言处理(natural language processing) 计算机科学的重要分支之一. 是用计算机处理自然语言信息的过程和有关技术. 自然语言指的是人类自身的书面或口头形式的语言，如汉语、英语、日语等，它是相对于人造的形式化的计算机语言而言的. 处理自然语言的关键是要让计算机理解自然语言，因此，自然语言处理通常又称为自然语言理解. 用计算机技术处理自然语言是非常困难的，需要语言学家、计算机科学家、心理学家、逻辑学家、思维科学家们共同研究来协同攻关. 它的研究在实用方面和理论方面都有重要意义. 在实用方面，使用自然语言同计算机通信的技术，用户可以方便地使用数据库、知识库、软件包、专家系统等. 在理论方面，自然语言处理居于人工智能研究的核心地位，它的最终目标之一是要搞清楚人类理解语言的机制，这对于智能的研究极有意义. 当前自然语言处理的应用领域主要有以下几方面：

1. 自然语言人机接口. 指各种知识处理系统与人类的接口，可具有简单句理解、复杂句理解、篇章理解等不同功能.

2. 文摘生成. 机器有能力产生与输入文本长短不同的摘要.

3. 机器翻译. 用计算机把一种自然语言翻译成另一种自然语言.

4. 问答系统. 机器理解一种自然语言表达的提问，然后用同一种自然语言来回答问题.

自然语言理解(natural language understanding) 计算机科学的重要分支之一. 是计算机能够理解人们用口头谈话方式或文字书写方式所表达的语言. 它是人工智能领域内的中心课题之一，其研究对象是如何使用自然语言同计算机通信的技术. 语言理解过程实质上是如何把一种语言表达(源表达)转换为另一种语言表达(目标表达)的转换过程. 要完成这种转换过程，不仅要有语言学方面的知识，而且还要有与论题有关的客观世界知识及推理知识等. 而要让计算机学会这些知识是很困难的. 计算机理解自然语言有两个方面：书面理解；语音识别、理解与合成.

书面理解是把文字输入计算机，使计算机能够读懂，并且用文字做出回答. 一个简单句的理解通常包括词汇分析、句法分析、语义分析和语用分析等过程.

1. 词汇分析. 把有序的、离散的线性字符串进行"切词"和词义判断. 切词通常依靠词典与知识推理相结合的方法，同时采用正向扫描与逆向扫描的一致性来校正. 词义判断是将词典中的每一词条与它们所在的上下文中的信息联系起来，以确定每个词汇的正确含义.

2. 句法分析. 确定输入的单词序列能否构成合法的句子，并抽出合法句子的结构. 这要依赖于所采用的语法规则. 通常采用的语法有上下文无关语法、转换语法、扩充转移网络语法等.

3. 语义分析. 赋予句法结构以语义解释，以识别一句话所表达的实际含义. 常用的语义分析技术有语义语法、格语法、语义网络、概念从属等.

4. 语用分析. 为确定句子的真正含义，根据语用环境的客观世界知识对表达的结构重新做出解释.

语音识别、理解与合成，指用口语对机器讲话，使机器能够听懂. 建立一个理解口语的系统，除要有书面理解系统所具有的功能外，还要有足够的知识来处理出现在语言信号中的噪音及多义表达. 它首先要分析各种辅音、元音、声调和语调等语音成分的物理特征，把这些不同的特征转换成数码形式供机器辨认. 还要研究音位和音位变体之间的关系，以及

从连续的音流中截取单词、音节，识别各种连续音变、人和人之间的语音差别等.

自然语言理解可以不涉及合成目标语言，这是与自然语言处理的区别.

机器翻译(machine translation)　一种计算机应用学科.指应用计算机技术把一种自然语言(源语言)翻译成另一种自然语言(目标语言).是一门涉及语言学、计算语言学和计算机科学等学科的边缘性应用学科.

机器翻译可分为文字翻译和语音翻译.按照源语言和目标语言的数量，又可分为一对一、一对多、多对一和多对多语言的机器翻译.机器翻译系统的实现方式大致可分为直接方式、转换方式和中间语言方式等三种.直接方式从源语言的表层句子出发，直接生成目标语言.转换方式先将分析源语言的表层句子所得到的内部结构变换为目标语言的内部结构，然后生成目标语言的表层句子.中间语言方式先将源语言变换成为一种不依赖于特定的自然语言的中间语言形式，然后由中间语言直接生成目标语言.这种方式适合于多语种互译的机器翻译系统，但中间语言的创建工作非常困难.当代多数实用型机器翻译系统基本上采用转换方式.它们通常包括词典、句法分析、语义分析、转换和综合等四个子系统.

1.词典子系统.通过自动查找机器词典，提供源语言句子中每个词的有关信息.机器词典除给出每个源语言句子中每个词对应的目标语言的译法外，还给出翻译过程中需要使用的其他一系列信息，例如词类、词组、语法、语义和目标语言综合等信息.

2.句法分析子系统.对源语言句子进行句法分析，得出源语言句子的句法结构描述(例如句法树等).

3.语义分析子系统.进行语义分析的目的是获得源语言句子的语义表示或句法分析中可能遇到的歧义问题.

4.转换和综合子系统.根据求得的词汇信息、句法信息和语义信息，参考源语言与目标语言之间在词汇和句法等方面的差异，找出对应的映射规则，进行必要的词汇转换和句法结构转换，最后综合出合乎目标语言规范的正确译文.

机器翻译的特点是速度快，节省翻译人员的工作量；但目前翻译质量尚不理想.

人工神经网络(artificial neural network)　一种新型的信息处理系统.是试图以模拟人脑神经系统的组织方式来构成的非线性自适应信息处理系统，具有学习和自我组织能力，是由大量处理单元(神经元)在系统内部广泛互连而成的可训练网络.它是在现代神经科学研究成果的基础上提出的能模拟人的智能和形象思维能力的一条重要途径和方法.在人工神经网络中，网络的信息处理由神经元之间的相互作用来实现，知识与信息的存储表现为网络元件互连之间的分布式物理连结，网络的学习和识别决定于各神经元连结权系数的动态演化过程.与传统的以逻辑符号为基础的符号主义人工智能的行为和能力不同，连结主义的神经网络具有并行分布式信息处理、容错、学习、自组织和实时处理等能力.有利于处理感知功能、直觉顿悟、形象思维、自组织自学习等.基于神经器件和神经网络的神经计算机具有速度快、并行处理、自学习、自组织、分布式存储、能模拟人类感知和智能活动等功能，已在军事、工业、管理等部门逐渐得到应用.

近年来，人工神经网络与模糊系统、遗传算法、进化机制等结合，形成计算智能，成为人工智能的一个重要方向，将在实际应用中得到发展.

机器人(robot)　机械电子一体化的高级自动装置.它具有类似某些生物器官功能，使用程序控制来完成特定操作或移动任务.现代机器人有两个特点：一是有类似于人的上肢的功能，可做空间多自由度运动；二是使用计算机控制，能灵活适应工作情况的变化.机器人的发展同计算机的发展有密切关系.虽然20世纪50年代出现了现代机器人的原型，20世纪60年代已有了实用的机器人，但发展得很慢；直到20世纪70年代出现微型计算机后，机器人才开始蓬勃地发展.

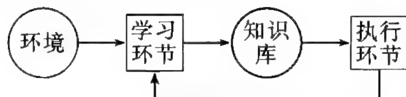
机器人的发展大致可分为三个阶段：第一阶段是固定程序和遥控式机器人；第二阶段是可变程序和示教再现式机器人；第三阶段是智能机器人.机器人是离散型生产过程自动化的必然产物，特别是在危险有害、单调、孤寂狭小的环境下，由机器人工作更显得优越.应用较多的是产业机器人，用于各种制造业.此外，还应用于种植业、采掘业、建筑业、商业、服务业以及家务劳动等方面.还应用于军事上，做排雷、流动哨兵、操纵武器等工作.

智能机器人(intelligent robot)　具有人类所特有的某种智能行为的机器人.是机器人发展的高级阶段，具有感觉、认识、判断、学习等功能，能自动识别周围环境并自动做出行动规划.在识别环境时，它主要通过视觉、听觉和触觉的敏感元件将信息输入到计算机，经过处理后，提取关键信息，并对它进行识别和理解.视觉是由电视摄像机输入二维图象，然后进行图象处理和识别.触觉包括力觉和滑觉，用于调节机械手的抓握力，使之恰好能抓住物体，既不用力过猛，又不致滑落.听觉可作为人机联系的手段.在有限范围内，机器人可以接受自然语言.为了制定行动规划，机器人应具有自动求解问题的能力.例如，可用一套特殊的产生式规则在状态空间中搜索求解.为了得到操作序列，可以从当前的状态集出

发,进行正向搜索,也可以从目标状态集出发进行逆向搜索,也可根据目标状态和当前状态的差选择合适的操作(手段-目的分析法)等.

机器学习(machine learning)　亦称计算机自动获取知识.人工智能学科的一个分支.研究计算机怎样模拟或实现人类的学习行为,以获取新的知识和技能,重新组织已有的知识结构,使计算机系统不断改善自身的智能行为.

对于机器学习的准确含义,存在着一些不同的观点.其中西蒙(Simon,H.A.)提出,学习是系统用于改进性能的过程,在此过程中,系统用新方法和新知识去改进现有方法和知识,以改善系统自身的性能.米哈尔斯基(Michalski,R.S.)认为,学习是构造或修改对于所经历事物的表示.从事专家系统研制的人们则认为学习是计算和自动获取知识.按照第一种观点,可以给出如图所示的学习系统的简单模型.图中,圆圈内是信息的集合,方块内是过程,箭头表示数据流向.环境把某些信息提供给学习环节,学习环节利用这些信息来改进知识库,执行环节利用知识库去完成任务,在完成任务过程中的信息可以作为对学习环节的反馈.按照环境提供给学习环节



一个学习系统的简单模型

的信息的水平,可以把学习方法分成以下五类:

1.由记忆学习,即存储新知识.学习环节不改变环境提供的信息的水平和形式,只需把这些信息加入知识库中.

2.由告知学习,即采纳建议.环境提供的信息过于抽象或一般化,它的水平高于执行环节所用信息的水平.因此,学习环节必须进行称为实用化的转换,由告知的信息推出结论、做出假设、填补细节和确定何时要求进一步的建议.

3.由实例学习,即归纳.这时,环境提供的信息过于特殊和具体,它的水平低于执行环节所用信息的水平.因此,学习环节必须由特殊的实例归纳出一般的规则.

4.由类比学习.是通过把十分类似于所需的新概念或技巧的已有知识转换或扩大为适用于新情况的形式来获取新方法或知识.

5.由观察和发现学习,又称“不受监督的学习”.它包括发现系统、理论形成任务、产生分类准则和类似的没有外部教师参与的学习.系统从环境得到的信息也是低水平的实例,但这些实例没有划分为正例和反例,因此,要求学习环节使用比其他方法更多的推理.

目前,具备自动学习功能的机器学习方法主要有归纳学习、分析学习、遗传学习、连结学习等.归纳学习从具体事例出发,通过归纳推理得到新的概念或知识.分析学习是利用背景或领域知识,通过对典型实例的分析、演绎推导,形成新的知识.遗传学习模拟生物繁殖中的遗传变异原则和达尔文的自然选择原则,通过某种目标函数,决定各种变体在发生突变和重组后是被淘汰还是继续生存.连结学习是在人工神经网络中,通过样本训练,修改神经元间的连结强度,甚至神经网络本身的结构的一种学习方法.但这一领域正引起越来越多的关注,也必将取得更快的进展.

模式识别(pattern recognition)　识别模式信息的一种过程.是利用计算机或其他装置对物体、图象、声音、文字等模式信息进行自动鉴别、分类和解释的过程.模式信息一般是连续的,可以通过抽样或数字化变为数字信息.常见的模式识别方法有统计模式识别法和句法模式识别法两类.它们都首先进行输入预处理,把识别对象抽象成特征空间中的一个点或描述语言中的一个语句,然后再用统计模式法或句法识别法进行鉴别和分类.这两种方法的选择取决于问题的性质.如果被识别的对象极为复杂,而且包含丰富的结构信息,一般采用句法模式识别法;被识别对象不很复杂或不含明显的结构信息,一般采用统计模式识别法.这两种方法不能截然分开,在句法模式识别法中的基元本身就是用统计模式识别法抽取的.在应用中,把这两种方法结合起来分别施加于不同的层次,常能收到较好的效果.模式识别可用于文字和语音识别、遥感图象识别和医学诊断等方面.模式识别与人工智能、图象处理的研究有交叉关系,它的研究成果已应用于人工智能研究的景物理解、自然语言理解、图象处理中的图象分析等方面.

统计模式识别法(statistical approach to pattern recognition)　亦称决策理论方法.识别模式信息的一种方法.是发展较早也比较成熟的一种模式识别方法.这种方法的识别过程大致可分为以下几个步骤:

1.数字化.把被识别的对象变换为适于计算机处理的数字信息.

2.预处理.除去混入的干扰信息并减少某些变形和失真.

3.特征抽取.从输入模式中抽取一组特征,即将输入模式从对象空间映射到特征空间,这时的模式可用特征空间中的一个点或一个特征向量表示.

4.模式鉴别和分类.即从特征空间再映射到决策空间,为此而引入判别函数,由特征向量计算出相应于各类别的判别函数值,通过判别函数值的比较

实行分类.

根据判别函数中参量的表示方式的不同,统计模式识别法可分为参量决策法和非参量决策法.前者的判别函数是概率函数,参量是概率参量,特征向量是随机向量,后者的判别函数则采用非概率函数.统计模式识别法一般用于识别对象不很复杂或不含明显的结构信息的情况.

决策理论方法(decision theory method) 即"统计模式识别法".

句法模式识别法(syntactic approach to pattern recognition) 亦称结构方法或语言学方法.模式识别的一种方法.把被识别的模式(样本或图形)按其结构组合成一定的语句,然后用句法模式识别法确定其属于哪一个类别.这种方法把一个模式描述为较简单的子模式的组合,子模式又可描述为更简单的子模式的组合,最后得到一个树形的结构描述,在底层的最简单的子模式称为模式基元.基元代表模式的基本特征,不应含有重要的结构信息.模式这种以一组基元和它们的组合关系来描述的形式,称为模式描述语句,这相当于在语言中,句子和短语用词组合,词用字符组合一样.基元按照语法规则组合成模式.因此,一旦基元被鉴别,识别过程即可按句法分析进行,即分析给定的模式语句是否符合指定的语法.满足某类语法的模式即被分入该类.这一过程称为模式分类.句法模式识别法可用很小的语言集合去描述数量很大很复杂的图形集合.这种方法可用于识别包含丰富的结构信息的极为复杂的对象.

语言学方法(linguistics method) 即"句法模式识别法".

归结反演(resolution refutation) 亦称消解反演.它是一种求解问题、证明定理的计算机过程.所使用的证明方法与数学中的反证法思想十分相似.归结反演的问题描述及具体证明过程是给出一个公式集 S 和目标公式 L,通过反证或反演来求证目标公式 L,其步骤为:

1. 否定 L,得 $\sim L$.
2. 把 $\sim L$ 添加到 S 中去.
3. 把新产生的集合 $\{\sim L, S\}$ 化成子句集.
4. 应用归结原理(参见"归结原理"),力图推导出一个表示矛盾的空子句.

规则演绎系统(rule-based deduction system) 亦称基于规则的演绎系统.它是一种求解问题、证明定理的计算机系统,而且是一种采用直接证明法思想(相对于归结反演方式的反证法思想而言)的计算机系统.此种系统又细分为规则正向演绎系统、规则逆向演绎系统以及规则双向演绎系统三种.在规则演绎系统中,总将被求解的问题描述为:事实、规则和目标.规则正向演绎系统从事实出发,通过使用规则,试图推导出所需目标.规则逆向演绎系统则从目标出发,通过使用规则,试图推证出所需事实都存在.规则双向演绎系统则结合使用正向与逆向两种系统求解技术来求解问题、证明定理.规则演绎系统对被求解问题中的事实、规则和目标都有一定的限制,也即都有其使用条件.

1. 规则正向演绎系统的使用条件:

1) 事实表达式为任意谓词公式.

2) 规则形式为 $L \to W$(其中 L 为单文字,W 为任意谓词公式).

3) 目标公式为文字析取式.

2. 规则逆向演绎系统的使用条件:

1) 事实表达式为文字合取式.

2) 规则形式为 $W \to L$(其中 L 为单文字,W 为任意谓词公式).

3) 目标公式为任意谓词公式.

启发式搜索(heuristic search) 一种计算机搜索策略.在状态空间问题求解过程中,如何找到从初始状态到目标状态的一条"解路径",存在着许多种搜索策略.从大的方面可分为盲目搜索和启发式搜索两大类.盲目搜索也称无启发信息的搜索,它或者按照事先规定的搜索路线去进行搜索,或者是按已经付出的代价多少来决定下一步要搜索的结点,如宽度优先搜索、深度优先搜索以及等代价搜索等都属于盲目搜索.宽度优先搜索是按"层"逐步进行搜索的;深度优先搜索是按纵深方向逐步进行搜索的;而等代价搜索则按到达当前搜索树各"端叶点"目前已付出的代价多少来决定下一步要搜索的结点.盲目搜索没有利用问题本身的特征信息,在决定要被扩展的结点时,没有考虑该结点在解的路径上的可能性有多大,以及求出的解是否为最优解等.因此,这种搜索方法具有较大的盲目性,产生的无用结点较多,搜索空间较大,效率不高.启发式搜索要用到问题自身的某些特征信息,特别是经验的启发性信息,以指导搜索朝着最有希望的方向前进,从而达到减少搜索范围、降低问题复杂度的目的.启发性信息按其用途可分为以下三种:

1. 用于决定要扩展的下一个结点,以免像在宽度优先搜索或深度优先搜索中那样盲目地扩展.

2. 在扩展一个结点的过程中,用于决定要生成哪一个或哪几个后续结点,以免盲目地生成所有可能的结点.

3. 用于决定某些应该从搜索树中抛弃或修剪的结点.

对于上述第一种启发性信息,通常使用与问题相关的估价函数来实现.估价函数能够提供一个评定候选扩展结点的方法,以便确定哪个结点最有可能处在通向目标的最佳路径上.一种估量某个结点

"希望"程度的重要方法是对各个结点使用估价函数值.如用符号 f 来标记估价函数,用 $f(n)$ 表示结点 n 的估价函数值.一般可通过如下方式来建立这种实值估价函数:试图确定任意结点处于最佳路径上的概率;提出任意结点与目标集之间的距离量度或差别量度;或者在棋盘式的博弈和难题中根据棋局的某些特点来决定棋局的得分数,如此等.

A^*算法(algorithm A^*)　一种计算机算法.用于状态空间问题的求解过程.A^*算法使用了一个特别的估价函数 f,对状态空间中的任一个节点 n 来说,

$$f(n)=g(n)+h(n),$$

其中的 $g(n)$ 是对 $g^*(n)$ 的估计,且 $g(n)>0$;$h(n)$ 是 $h^*(n)$ 的下界,即对所有的 n 均有

$$h(n)\leqslant h^*(n).$$

而上述的 $g^*(n)$ 表示从始点 S 到达该结点 n 的最小代价路径的代价,而 $h^*(n)$ 表示从该结点 n 到达某个目标结点 G 的最小代价路径的代价.也就是说 $f(n)$ 是约束通过结点 n 的一条最小代价路径之代价的一个估计.A^*算法总扩展当前具有最小 f 值的那一个结点 n. A^*算法具有如下一些重要特征:

1. A^*算法是可纳的.对任一个可解状态空间图(即从初始结点到目标结点有路径存在的图)来说,A^*算法一定能够在有限步内终止,且在终止时找到了最优解路径.

2. A^*算法的最优性取决于下界函数 $h(n)$ 与 $h^*(n)$ 函数的接近程度.即在满足 $h(n)\leqslant h^*(n)$ 的前提下,$h(n)$ 的值越大越好.$h(n)$ 的值越大,表明它携带的启发性信息越多,搜索时扩展的结点数也就越少,搜索效率就越高.具体地说,有如下结论:设 f_1 与 f_2 是对同一问题的两个估价函数,

$$f_1(n)=g_1(n)+h_1(n),$$
$$f_2(n)=g_2(n)+h_2(n),$$

A_1^* 与 A_2^* 分别是以 $f_1(n)$ 及 $f_2(n)$ 作为估价函数的 A^*算法,并且设对所有非目标结点 n 均有 $h_1(n)<h_2(n)$,则 A_1^* 所扩展的结点数决不会比 A_2^* 扩展的结点数少,即 A_2^* 扩展的结点集是 A_1^* 扩展的结点集的子集.

自动程序设计(automatic programming)　一种计算机程序系统.该系统或者用来实现程序综合,或者用于完成程序正确性验证.程序综合用于自动编程,即用户只需告诉计算机"要做什么",而无需说明"怎样做",计算机就可自动实现程序的设计.具体地说,首先给系统输入关于所设计的程序要求实现某个目标的高级描述,然后系统对这种高级描述进行处理,通过规划过程,自动生成一个能完成所给目标的具体程序.程序正确性验证是要得出一套理论和方法,通过运用这套理论和方法就可证明程序的正确性.目前常用的方法是用一组已知结果的数据对程序进行测试,如果程序的运行结果与已知结果一致,就认为程序是正确的.但此种方法仅对简单程序有效.程序正确性验证至今仍是一个比较困难的课题,有待进一步研究解决.自动程序设计所涉及的基本问题与定理证明和机器学习有关,它也是软件工程和人工智能相结合的课题.

人工智能程序设计语言(programming language of artificial intelligence)　亦称人工智能语言.一类计算机程序设计语言.使用这类语言,可以方便而有效地建立人工智能系统.下面列出的是最常用的几种人工智能程序设计语言,简单介绍如下:

1. IPL.一种早期的表处理语言.IPL 是 Information Processing Language 的首字母缩写.该语言是专为人工智能应用而设计的早期语言之一.表处理是 IPL 的主要目标.现在它已完全被 LISP 这样的高级语言所取代.但其意义犹存,因为它是研制人工智能语言的起点.

2. LISP.目前应用最广泛的表处理语言.LISP 是 List Processing(表处理)的首字母缩写.该语言是最早和最重要的逻辑型编程语言之一.它于 1958 年由美国的麦卡锡(McCarthy,J.)提出,并于 1960 年发表了他的第一篇关于 LISP 的论文.之后,LISP 语言很快得到广泛的应用,到目前为止,仍有使用 LISP 语言来编写人工智能系统的.LISP 语言具有各种不同的版本.使用最广泛的是 INTERLISP,MACLISP 和 COMMON LISP.其中 COMMON LISP 是在 20 世纪 80 年代初,由美国二十几家大学和研究所联合设计的,它是 LISP 语言发展的新的里程碑.

3. SAIL. SAIL 是 ALGOL 语言的变种,它最类似于常规的通用程序设计语言.但具有支持相关存储器、能直接与汇编语言过程相连结等附加特性.由于 SAIL 具备一门程序设计语言的所有标准特性,所以常用来求解涉及大量传统计算的人工智能问题.

4. PLANNER.一种便于目标定向处理的早期语言.PLANNER 是建立在 LISP 的基础之上,为问题求解及定理证明而提出的一种语言,该语言能实现传统的正向推理和面向目标的逆向推理等功能.一个 PLANNER 程序包括断言和定理两部分,断言部分用于指出已知事实,而定理部分则描述如何根据旧事实推出新事实.它们分别被存储在断言数据库和定理数据库中.

5. KRL. KRL 是 Knowledge Representation Language 的首字母缩写.该语言建立在 LISP 之上,易于表达框架结构中的知识,是一种能够支持复杂

框架结构的语言.

6. PROLOG. 目前应用最广泛的逻辑程序设计语言. PROLOG 是 Programming in Logic 的首字母缩写. 该语言是一种基于一阶谓词逻辑的产生式规则语言. 它由法国的考莫瑞尔(Colmerauer, A.)和他在马塞大学的助手于 1972 年发明的. 该语言的一个最显著的特点是想让程序员能以接近自然语言的形式,告诉计算机做什么,而不必像传统的程序设计语言那样,必须告诉计算机怎样做.

PROLOG 描述功能强,具有逻辑推理能力,语法简单,使用方便,已在专家系统、定理证明系统、数据库和知识库系统、机器人规划与控制系统、自然语言理解、计算机辅助设计等诸多领域得到成功的应用. PROLOG 程序一般可分为事实和规则两部分. 事实用只有句首的 horn 子句表示,规则由一个完整的 horn 子句表示. 而 PROLOG 程序的执行则通过使用只有句体的 horn 子句并以询问的方式来实现. 用 PROLOG 语言编程包括:规定操作对象以及关系的一批事实;规定操作对象以及关系的一批规则;询问操作对象以及关系的具体问题. PROLOG 本身就是一个使用归结法的演绎推理机,它通过合一、置换、消解、回溯和匹配等机制来求解问题. PROLOG 语言也有许多不同的版本,如 PROLOG, C-PROLOG, H-PROLOG, micro-PROLOG 以及 Turbo-PROLOG 等.

智能控制(intelligent control) 简称 IC. 一种计算机控制过程. 它是无需人的干预就能够自主地驱动智能机器实现其目标的控制过程. 这是用机器模拟人类智能的又一重要领域. 1985 年 8 月, IEEE 在美国纽约召开了第一届智能控制学术讨论会. 1987 年 1 月,在美国费城由 IEEE 控制系统学会与计算机学会联合召开了第一次智能控制国际会议. 智能控制具有明显的跨学科的多元结构特点. 现有的智能控制结构理论有二元、三元和四元交集结构等. 它们分别由下列各交集表示:

$$IC = AI \cap AC;$$

$$IC = AI \cap CT \cap OR;$$

$$IC = AI \cap CT \cap IT \cap OR;$$

式中,各子集的含义如下:

IC:智能控制;

AI:人工智能;

AC:自动控制;

CT:控制论;

OR:运筹学;

IT:信息论.

智能控制的研究领域包括智能机器人控制、智能调度与规划、专家控制系统、语音控制、康复智能控制器和智能仪器等. 智能控制具有下列特点:

1. 同时具有以知识表示的非数学广义模型和以数学模型表示的混合控制过程.

2. 智能控制的核心在高层控制,即组织级.

3. 智能控制是一门边缘交叉学科.

4. 智能控制是一个新兴的研究领域.

现已开发出许多智能控制理论与技术,并将它们应用于具体的控制系统中,如分级控制理论、分级控制器设计的熵方法以及智能逐级增高而精度逐级降低原理等. 另外还有一些针对控制系统应用的理论研究,如自学习与自组织系统、神经网络、基于知识的系统、语言学和认知控制器等.

智能决策支持系统(intelligent decision support system) 简称 IDSS. 一种重要的计算机决策系统. 它是传统决策支持系统(DSS)与人工智能技术相结合的产物,它在形成决策方案的过程中运用了知识以及知识处理的有关技术. 而传统的决策支持系统则是在管理信息系统(MIS)基础上发展起来的一种计算机管理系统,它是管理科学的一个分支. 智能决策支持系统是在 20 世纪 80 年代初提出来的. 由于它将决策支持系统与人工智能技术特别是专家系统技术相结合,所以它既保持有传统决策支持系统中数值分析的优势,又能充分发挥专家系统中知识及知识处理的特长,从而使系统既可以进行定量分析,又可以进行定性分析,能有效地解决半结构化及非结构化的问题,扩大了决策支持系统的应用范围,提高了系统求解问题的能力.

智能决策支持系统通常由数据库系统、模型库系统、方法库系统、人机接口系统及智能部件(也称知识库系统)等五部分所构成. 数据库系统由数据库及其相应的管理软件所构成,主要特征有:数据独立性,数据共享性,统一管理性,可修改及可扩充性,安全保密性以及最小冗余性. 模型库系统由模型库及其相应的管理软件所构成. 模型库是用于存储模型及其基本单元的,它具有共享性及动态性. 而模型库管理软件的主要功能为:创建模型、模型管理以及运行控制. 方法库系统由方法库及其相应的管理软件所构成,用于向系统提供通用的决策方法、优化方法以及软件工具等,并实现对方法的管理. 人机接口系统又称人机界面或用户界面系统. 人机接口系统要为决策者提供一个方便、友好的交互环境,能使用户充分了解系统的运行情况、运行结果及推理结论,并为决策者提供控制系统运行以及模型选择等方面的权利;另外还应提供在线帮助及保密等功能. 智能部件也称知识库系统,用于模拟人类决策过程中的某些智能行为. 由于它是面向决策支持的,除应具有特定的决策知识外,还应包含与模型、方法有关的各种知识.

机 器 证 明

机器证明(theorem proving by machine or automated deduction) 计算机科学的重要研究方向之一,也是人工智能领域的重要组成部分.原意是用机械的步骤判定数学命题的真假,现在一般指用计算机证明定理或推导公式.根据不同要求,可以提出一理一证、一类一证或一般通用的机器证明方法.其形式又可以采用计算机辅助证明或计算机自动证明.

机器证明是数学机械化研究的一项重要内容.在计算机出现以前,这一领域已经走过了漫长的道路.17世纪,莱布尼茨(Leibniz,G. W.)和笛卡儿(Descartes,R.)分别从建立通用证明器和引进坐标、化一切问题为代数问题这两个角度,考虑了逻辑的和代数的机器证明问题.尽管当时的努力不足以取得明显的进展,却提出了有重大意义的研究方向.事实上,今天的机器证明理论,主要是沿着这两个方向发展和完善起来的.其中,希尔伯特(Hilbert,D.)、塔尔斯基(Tarski,A.)、贺布兰德(Herbrand,J.)和鲁宾孙(Robinson,J. A.)分别做出了重要的贡献.中国数学家吴文俊,从20世纪70年代后期开始致力于数学机械化及定理机器证明的研究,提出了一整套有效的、被国内外公认的吴方法或吴理论.吴理论本质上是采用代数方法研究机器证明.这一理论不但丰富了机器证明的内容,而且还带动了其他研究方法的发展.在吴理论的影响下,近年来还出现了GB法、例证法、数值并行法和可读性证明法等.这些方法的不断发展与完善,无疑具有重要的意义.机器证明就其方法而言是一些能判定某些命题真假的机械化方法,即一些算法.算法的作用域是由若干具有确定形式的命题组成的集.也就是说,算法工作开始时输入要检验的命题的前提与结论或只输入前提.算法工作结束时的输出可能有下列几种:

1. 简单地断言命题成立或不成立,即输出命题的真值.

2. 宣布算法对所输入的命题无效.

3. 输出若干附加条件,并断言在这些附加条件下命题的结论成立.

4. 输出证明的步骤.

5. 输出由前提能推出的多种结论.

实际上,上述描述不能包罗机器证明的各式各样的描述与问题的提法,因为它是一个迅速发展而远未成熟的研究领域.

一理一证法(proving method for one theorem) 一种证明方法.是为证明一个定理而特别设计的方法.它是数学研究中使用的传统方法.为了判定某一命题的真假,有时可以提出一套切实可行但却需要进行大量运算的方案,这种方案往往只有在高速的计算机上才能在合理的时间内实现.一理一证法不是机器证明研究的主要努力方向,但针对某些特殊难题的研究工作是有用的.

一类一证法(proving method for one class of theorems) 一种证明方法.是为一类定理的证明而提出的方法.在计算机上实现这种方法具有重要的意义,因为它使人一劳永逸地原则上完成了某一领域中命题真假判断的研究工作.

一般证明法(theorem prover) 一种证明方法.直观上可以证明任何定理的方法为一般证明法,或为通用证明法、万能法.事实上,即使在算术领域,也不存在能判定一切命题的一般证明法.已经知道,存在一种万能法,可以在有限步内验证由谓词逻辑描述的任何定理.然而,这种方法的效率往往很低.

辅助证明法(auxiliary proving method) 亦称半自动证明法.一种证明方法.是用机械步骤完成定理证明中部分工作的方法.通常是人制订了证明的框架,而由计算机完成某些局部的细节推演.例如,有人设计了用计算机辅助证明四色定理的方案并进行了实施.

自动证明法(automatic proving method) 一种证明方法.是可以在计算机上完成定理证明全部工作的方法.

代数方法(algebraic method) 一种证明方法.是把证明过程转换为代数式之间的推导和计算的方法.用代数方法证明几何定理,首先是由笛卡儿(Descartes,R.)于1637年提出的.后来希尔伯特(Hilbert,D.)又向前推进了一步,把这种方法对某一类命题算法化,从而开创了真正能用机械步骤推出一批几何定理的局面.代数方法本质上属于一类一证的方法,它适用于假设和结论能够用代数恒等式或代数不等式表达的命题的判定.通常,用HS记假设条件,C记命题的结论,$f_1,f_2,\cdots,f_r$, g 记关于变元 $x_1,x_2,\cdots,x_n$ 的多项式.如果

$$
\begin{aligned}
\text{HS:}\quad & f_1(x_1,x_2,\cdots,x_n)=0,\\
& f_2(x_1,x_2,\cdots,x_n)=0,\\
& \vdots\\
& f_r(x_1,x_2,\cdots,x_n)=0,\\
\text{C:}\quad & g(x_1,x_2,\cdots,x_n)=0,
\end{aligned}
$$

则命题HS⇒C的判定问题已被解决.代数方法是定理机器证明领域中20世纪70年代至80年代获得最成功的方法.

逻辑方法(logic method) 一种证明方法.是将演绎推理机械化并用于定理证明的方法.莱布尼茨(Leibniz,G. W.)曾作过这方面的设想,并提出通用语言的方案.此后,布尔代数的发展,特别是量词的

引入，使得一阶谓词具有更强的描述能力. 计算机出现以后，用逻辑方法从事机器证明的研究取得了诸如归结原理的重要结果，并影响到新一代计算机的研制. 逻辑方法的进一步发展的重要问题是改进算法以降低复杂度、提高效率，使之在计算机上更好地实现.

归结原理(resolution principle) 计算机科学的一条原理. 鲁宾孙(Robinson, J. A.)于 1965 年给出的说明子句集 S 是不可满足的有效方法. 设 S 中包含有子句 $\sim P \vee Q_1$ 和 $P \vee Q_2$，则 $Q_1 \vee Q_2$ 称为它们的归结式. 得到归结式的过程称为归结. 在命题逻辑的情形下，归结较为容易. 在谓词逻辑的情形下，归结的进行往往需要做合一处理. 然而，只要 S 是不可满足的，就可以从 S 出发，归结出空子句，反之亦然，并且这一过程是有限的.

贺布兰德定理(Herbrand theorem) 有关一阶逻辑定理证明的重要定理之一. 将任意用一阶逻辑写出的合适公式 T 转换成析取式之后，都可以得到一个相应的子句集 S，并且 T 的真伪性判定就相同于 S 的不可满足性的判定. 1930 年，贺布兰德(Herbrand, J.)指出：合适公式 T 取真值当且仅当在相应的贺布兰德解释下，S 是不可满足的. 这一过程一般体现为树形搜索，并且在 S 是不可满足的情形下，搜索过程在有限步内终止.

塔尔斯基方法(Tarski method) 第一个初等代数和初等几何命题的判定算法. 1950 年，塔尔斯基(Tarski, A.)证明实闭域的初等代数问题是可判定的，由此推知，初等实几何是可机械化的. 用塔尔斯基方法，理论上可判定初等几何中任一命题的真伪. 但由于运算量太大，至今未能在计算机上证明稍难的定理. 塔尔斯基定理可表述为：设 $f_i(x_1, x_2, \cdots, x_n)$ 和 $g(y, x_1, x_2, \cdots, x_n)$ 是有序域 K 上的多元多项式，这里 $i=1,2,\cdots,m$，而 y 在 g 中真正出现，则有机械化方法可以判断. 对 K 的有序的代数封闭域 $\overline{K}$ 中的任意满足方程组

$$f_i(x_1, x_2, \cdots, x_n)=0 \quad (i=1,2,\cdots,m)$$

的数组 $a_1, a_2, \cdots, a_n$，方程 $g(y, x_1, x_2, \cdots, x_n)=0$ 在 $\overline{K}$ 中是否恒有解答.

机器发明(discovering by machine) 相伴于机器证明. 利用计算机得到新定理的过程. 其方法有三：

1. 用机器证明方法检验人的猜测. 例如，在一条二次曲线上有 6 个点 $A_1, A_2, \cdots, A_6$，取其中四个不同的点 A_i, A_j, A_k, A_l，作直线 A_iA_j 与 A_kA_l 之交点 $P_{ij,kl}$，称为巴斯卡点. 在由 $1,2,\cdots,6$ 生成的对称群中，取一个阶数为 6 的子群 G，例如取

$$G=\{1, g, g^2, g^3, g^4, g^5\} \quad (g=(1,2,3,4,5,6)).$$

以 G 中各元素 σ 作用于点 $P=P_{ij,kl}$，得点

$$\sigma P = P_{\sigma(i)\sigma(j),\sigma(k)\sigma(l)},$$

则 6 点 $P, gP, g^2P, g^3P, g^4P, g^5P$ 在一条非退化的二次曲线上. 从不同的 g 与 P 可得这样的二次曲线 60 个. 这是用吴文俊方法在计算机上发明的定理之一.

2. 用计算机搜索对象的未知性质. 例如用张景中等人提出的方法和通用程序 GISS，能自动搜索出几何图形的大量非平凡性质，并对每一性质给出传统方式的证明.

3. 用计算机推导未知的等式或不等式. 例如，用杨路提出的方法和通用程序 Inventer，能对给定的图形推出某些指定参数之间应满足的不等式.

欧几里得模式(Euclid mode) 初等几何的传统证明模式. 即从公理和定义出发，用演绎推理的方法证明一些定理；再从公理、定义和已证明的定理出发，证明新的定理. 证明的过程仅仅是推理过程，证明的方法则因命题不同、因人不同而各具巧思.

笛卡儿模式(Descartes mode) 亦称解析几何方法. 论证几何问题的一种模式. 引入坐标系，把几何命题的假设和结论转化为代数等式或不等式，从而把几何命题化为代数命题. 用这种方法判定几何命题真伪，判定过程也就是代数演算的过程，即符号计算的过程. 解析几何方法使命题的判定变得较为有章可循，但仍不能达到初等几何命题判定的机械化.

希尔伯特模式(Hilbert mode) 亦称希尔伯特机械化定理. 论证几何问题的一种模式. 它出现于希尔伯特(Hilbert, D.)的名著《几何基础》第六章之末. 原叙述为："设一种平面几何，其中公理 Ⅰ 1-3，Ⅱ，Ⅳ* 都满足，而且帕斯卡定理正确，则这几何中的每一条纯粹的交点定理，都可以通过作适当的辅助点和辅助直线表为有限个帕斯卡构形的组合". 其中"公理 Ⅰ 1-3"指希尔伯特公理系统中平面上的第 Ⅰ 类关联公理，"Ⅱ"指第 Ⅱ 类次序公理，"Ⅳ*"指狭义的第 Ⅳ 类平行公理. 所谓帕斯卡定理是指"若相交的两直线 l 与 l' 上各有互不相同的三点 A, B, C 与 A', B', C'，若

$$BC' \parallel B'C, \; AB' \parallel A'B,$$

则必有

$$AC' \parallel A'C".$$

所谓帕斯卡构形，则指相当于帕斯卡定理内容的一个图形. 所谓纯粹的交点定理，是指定理中只含有关于点和直线的位置关联及关于直线平行性的叙述，而不用其他关系(如合同和垂直). 吴文俊指出，希尔伯特的这条定理解决了平面几何中只涉及关联性的一类命题的机械化判定问题，因而是最早的机械化定理. 但希尔伯特本人并不一定意识到这一点.

机械化定理(mechanical theorem)　断言某类命题可机械化判定的定理. 关于一门学科的某一部分定理具有机器证明方法的断言如果成立,便称为相对于这一学科部分的机械化定理,也称为这一学科部分是可以机械化的,如欧氏无序几何是可以机械化的.

等式型命题(proposition about equalities)　一类形式命题. 是假设和结论部分都可以用多项式等式关系来表达的那类命题.

不等式型命题(proposition concerning inequalities)　与等式型命题相应的另一类命题. 假设部分由多项式等式或不等式关系来表达,而结论部分包含多项式不等式关系的那类命题. 但当变元为实数时,不等式 $f \geqslant 0$ 可以写成 $f = t^2$ 的形式,其中 t 是一个独立变元,所以假设部分可以由多项式等式组成.

多项式组(polynomial set)　一个基本概念. 指由有限个非零多项式构成的集合. 在几何定理机器证明中,命题常被转化为这种形式:给定的一个多项式组的多项式的公共零点是否都是另一个给定的多项式的零点. 多项式组常用记号 PS 表示.

多项式的主变元(main variable of polynomial)　多项式的一个变元. 通常指多元多项式中,具有最大下标的变元,有时为了叙述的方便,也可以临时指定某一变元为主变元. 引入主变元的好处,是便于把多元多项式表示成关于主变元的一元多项式的形式.

多项式的类(class of polynomial)　一个基本概念. 指多项式的下标,即多项式中按变量的顺序所含变量的最大下标.

多项式的秩(rank of polynomial)　对多项式进行排序的一种标志. 若多项式的类为 p,则其秩由 p 和关于 x_p 的次数确定. 设多项式 f 和 g 的类分别为 p_f 和 p_g,而关于相应变元的次数分别为 d_f 和 d_g,如果 $p_f > p_g$,或者 $p_f = p_g$,而 $d_f > d_g$,则称 f 的秩不低于 g 的秩.

多项式的初式(initial term of polynomial)　亦称多项式的首项系数. 多项式关于其主变元的最高次项的系数. 设多项式 f 的类为 p,关于主变元 x_p 的最高次项的系数即为 f 的初式. 它一般是关于变元 $x_1, x_2, \cdots, x_{p-1}$ 的多项式. 这些变元中可能有些不在初式中出现.

多项式的约化(reduction of polynomial)　一种多项式处理的方法. 是利用一个或几个多项式化简另一个多项式的一种基本方法. 设多项式 f 的类为 p,若多项式 g 关于 x_p 的次数低于 f,则称 g 已对 f 约化. 否则,可把 f, g 都看成 x_p 的单变元多项式,用 f 除以 g 得余式 r,则 r 已对 f 约化. 这一过程称为 g 对 f 的约化. 若 PS 是一组多项式,而 g 对 PS 中的每一个多项式都已约化,则称 g 对多项式组 PS 是约化的.

准升列(quasi ascending chain)　亦称三角化多项式组. 一组多项式. 若其中任两个多项式的类均不相同,则称该多项式组为准升列. 适当改变变元的名称和下标,总可以把一个准升列写成如下形式:

$$\begin{aligned}&f_1(u_1, u_2, \cdots, u_d, x_1),\\&f_2(u_1, u_2, \cdots, u_d, x_1, x_2),\\&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\&f_{s-1}(u_1, u_2, \cdots, u_d, x_1, x_2, \cdots, x_{s-1}),\\&f_s(u_1, u_2, \cdots, u_d, x_1, x_2, \cdots, x_{s-1}, x_s).\end{aligned}$$

准升列比一般的多项式组研究起来方便得多. 在上述准升列中,变元 u_i 称为参数或独立变元,变元 x_j 称为约束变元.

升列(ascending chain)　一种标准形式的准升列. 设 $f_1, f_2, \cdots, f_s$ 是准升列,如果每个 $f_i (i = 2, 3, \cdots, s)$ 对于 $f_1, f_2, \cdots, f_{i-1}$ 均已约化,则称此准升列为升列(参见"准升列").

升列的秩(rank of ascending chain)　对升列排序的标志. 升列的秩是由其所含多项式的秩决定的. 设有升列 A: $A_1, A_2, \cdots, A_t$ 和 B: $B_1, B_2, \cdots, B_s$,如果 A_r 的秩大于 B_r 的秩,而对 $i = 1, 2, \cdots, r-1$ 有 A_i 的秩等于 B_i 的秩,或者对 $i = 1, 2, \cdots, t$ 有 A_i 的秩等于 B_i 的秩,而 $s > t$,则称 A 的秩不低于 B 的秩.

多项式组的升列(ascending chain of polynomial set)　一种升列. 是多项式组 PS 中的若干多项式组成的升列.

极小升列(minimum ascending chain)　亦称基列. 一种升列. 一个多项式组的有限个升列中,秩不高于其他升列者,称为这些升列中的极小升列.

里特-吴整序原理(Ritt-Wu well ordering principle)　机器证明的吴文俊方法的一项重要内容. 里特-吴整序原理指出一种机械方法,使得从一给定的多项式组 PS 出发,能够得到若干升列 CS_i $(i=1, 2, \cdots, k)$,使得 PS 的零点集有下列形式的分解:

$$\mathrm{Zero}(\mathrm{PS}) = \bigcup_{i=1}^{k} \mathrm{Zero}(CS_i / J_i),$$

其中 Zero(PS)表示 PS 中多项式的公共零点集,J_i 是 CS_i 中全部初式之积,而 $\mathrm{Zero}(CS_i/J_i)$ 表示使 CS_i 等于零,同时又满足 J_i 不等于零的点的集合,即

$$\mathrm{Zero}(CS_i/J_i) = \mathrm{Zero}(CS_i) \backslash \mathrm{Zero}(J_i).$$

上述 Zero(PS)的分解,称为零点结构定理. 关于整序步骤可参见"多项式组的特征列".

余式公式(remainder formula)　机器证明的吴文俊方法的基本公式之一. 设有一升列 A: $A_1, A_2, \cdots, A_t$ 和一多项式 g,将 g 对 A 约化,即依次做除法:

$$I_t^{S_t} \cdot g = Q_t \cdot A_t + R_{t-1},$$

$$I_{t-1}^{S_{t-1}} \cdot R_{t-1} = Q_{t-1} \cdot A_{t-1} + R_{t-2},$$
$$\cdots\cdots$$
$$I_2^{S_2} \cdot R_2 = Q_2 \cdot A_2 + R_1,$$
$$I_1^{S_1} \cdot R_1 = Q_1 \cdot A_1 + R_0,$$

其中每次除法均把各多项式看成是以 A_i 的主变元为变元的一元多项式. 得到的 R_0 由 g 和 A 惟一确定,并称 R_0 为 g 对升列 A 的余式. 而 I_j 是 A_j 的初式,S_j 是某些非负整数. 则有余式公式

$$I_1^{S_1} \cdot I_2^{S_2} \cdot \cdots \cdot I_t^{S_t} \cdot g = P_1A_1 + P_2A_2 + \cdots + P_tA_t + R_0,$$

其中 $P_1, P_2, \cdots, P_t$ 是某些多项式. 由余式公式可见,当 $R_0 \equiv 0$ 时,A 的任一零点若不是 $I_1, I_2, \cdots, I_t$ 的零点,则必为 g 的零点.

多项式组的特征列(characteristic chain of polynomials) 多项式组的标准形式之一. 从一多项式组 PS 出发,找出它的一个极小基列 A_1,将 PS 中的 A_1 之外的多项式对升列 A_1 约化而求其余式,把诸余式添入 PS 得多项式组 PS_1. PS_1 与 PS 有相同的零点集. 再取 PS_1 的一个基列 A_2,重复上述步骤,依次得 $PS_2, PS_3, \cdots$,若 PS_k 有一基列 A_k,使得 A_k 以外的多项式对于 A_k 的余式均为零,则称 A_k 为 PS 的特征列. 里特-吴整序原理指出,按上述步骤进行有限次后,或得到一个特征列,或得到一个与 PS 有相同零点集的仅含参变量的多项式(参见"里特-吴整序原理").

不可约升列(irreducible ascending chain) 具有最简单结构的一种升列. 设 φ: $A_1, A_2, \cdots, A_n$ 是一升列,其中 A_i 是变元 $u_1, u_2, \cdots, u_d, x_1, x_2, \cdots, x_i$ 的多项式,系数在域 K 之中. 如果每个 f_i 在多项式环

$$\frac{K(u_1, u_2, \cdots, u_d)[x_1, x_2, \cdots, x_i]}{(f_1, f_2, \cdots, f_{i-1})}$$

中不可约,则称 φ 为不可约升列.

吴文俊方法(Wu Wenjun method) 一种定理机器证明方法. 是由吴文俊创立并加以系统化和完善的定理机器证明方法,简称吴方法. 吴方法以构造性的代数几何理论为工具,将几何定理的求证问题,经坐标化和代数化之后,转变成判定两个代数簇的有条件的包含关系,再加以解决. 应用吴方法建立的程序系统已经证明了数百条高难度的几何定理,同时还发明了若干新定理. 吴方法还可用于证明非欧氏几何中的定理,并能证明若干微分几何定理,及解决微分方程中的某些定性问题.

用吴方法判定一个命题,其本质是判定一组多项式的公共零点集是否被包含于另一多项式的零点集的问题. 分三步进行:第一步是把所给命题化为代数形式. 第二步是整序,即把刻画命题条件的多项式组 PS 经整序化为升列 AS. 第三步是求余,即将刻画命题结论的多项式 g 对升列 AS 约化求余式 R. 若 R 为 0,即可断定命题在非退化条件 $I_1I_2\cdots I_k \neq 0$ 之下成立,或者说命题一般成立. 其中 $I_1I_2\cdots I_k$ 是升列 AS 中各多项式的初式之积. 若 R 不为 0,则当 AS 为不可约升列时,可断定命题不真.

格若勃基(Groebner basis) 解决多项式理想理论有关问题的一个重要工具. 考虑域 K 上的以 $y_1, y_2, \cdots, y_n$ 为变元的多项式,并在两个单项式之间规定一种顺序:认为 $y_1^{s_1}, y_2^{s_2}, \cdots, y_n^{s_n} < y_1^{t_1}, y_2^{t_2}, \cdots, y_n^{t_n}$,如果有 $k \leqslant n$ 使 $s_k < t_k$,并对一切 $i < k$ 有 $s_i = t_i$,即字典序. 于是,每个非零多项式 p 可以惟一地写成

$$p = c_1m_1 + c_2m_2 + \cdots + c_km_k,$$

其中 c_k 是 K 中的非零元素,m_i 是诸 y_j 的方幂的积,并且 $m_1 > m_2 > \cdots > m_k$. 称 c_1m_1 为 p 的导项. 设有一多项式集合 S,g 是一个多项式,如果 S 中某多项式 f 的导项是 g 中某项 m 的因式,则称 g 是模 S 可约的. 设 $f = c_1m_1 + f_1$,并且 $m = sm_1$,将 g 中的项 m 换成 $-Sf_1/c_1$,得到 g_1. 称 g 的模 S 一步约化为 g_1. 显然 $g - g_1$ 属于 S 的理想. 重复这一过程,最后得到一个模 S 不可约化的多项式 g_k,称 g_k 为 g 的模 S 的一个正规形式. 一般地,g 的模 S 正规形式不惟一. 如果对任一多项式 g,g 的模 S 的正规形式是惟一的,则称 S 是它所生成的理想 $I(S)$ 的格若勃基. 若 S 是 I 的格若勃基,则 $g \in I$ 的充分必要条件是 g 的模 S 的正规形式为 0.

给出理想 I 的一个有限生成集 S,可以机械地构造出 I 的格若勃基. 算法如下:记 $S_0 = S$,并对 S_0 中任一对多项式

$$p_i = c_im_i + f_i \quad (i = 1, 2)$$

构造一新的多项式:设 m 是 m_1, m_2 的最小公倍数,而且 $m = s_1m_1 = s_2m_2$,则新的多项式

$$p = s_1f_1 - (c_1/c_2)s_2f_2.$$

把新构造的多项式模 S_0 的正规形式中非 0 者添入 S_0,得 S_1. 重复这一过程,得

$$S_0 \subset S_1 \subset S_2 \subset \cdots \subset S_k = S_{k+1},$$

则 S_k 是 I 的一个格若勃基. 将格若勃基中每一多项式对其余的多项式约化到模,其余多项式为正规形式,得到的一组多项式称为既约格若勃基. 对于给定的出发集 S 和指定的单项式之间的字典序,既约格若勃基是惟一的.

格若勃基方法(Groebner basis method) 一种定理机器证明方法. 是将格若勃基用于几何定理机器证明的方法. 格若勃基方法已用于几百条几何定理的证明,并能从给定前提条件导出多种有几何意义的结论,以发现新定理. 但在多数情形下,其效率不及吴方法. 设要判定的命题已化为代数形式,假设条件可表示为

$$h_i(u_1, u_2, \cdots, u_d, x_1, x_2, \cdots, x_r) = 0 \ (i = 1, 2, \cdots, n),$$

结论可表示为

$$g(u_1,u_2,\cdots,u_d,x_1,x_2,\cdots,x_r)=0.$$

格若勃基方法如下：

第一步：把 $x_1,x_2,\cdots,x_r$ 看成变元，计算$\{h_1,h_2,\cdots,h_n\}$的格若勃基 G，将 g 对 G 约化求出模 G 的正规形式. 若其正规形式为 0，则命题一般真（即在一组非退化条件下成立）. 若不为 0，则进行第二步.

第二步：引入新变量 z，计算$\{h_1,h_2,\cdots,h_n,zg-1\}$的格若勃基. 若此格若勃基为$\{1\}$，即仅含常数项——关于 u_i 的多项式，则命题一般真. 否则，命题不真.

在求取格若勃基的过程中，$u_1,u_2,\cdots,u_d$ 被看成参数，因而在格若勃基的表达式中，可能出现以 $u_1,u_2,\cdots,u_d$ 为变元的分式. 应用格若勃基证明定理时，要求这些分式的分母不为 0. 这也称为非退化条件.

结式(resultant)　一种行列式. 设有两个多项式

$$p(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0,$$
$$q(x)=b_kx^k+b_{k-1}x^{k-1}+\cdots+b_1x+b_0.$$

则行列式

$$\begin{vmatrix} a_n & a_{n-1} & \cdots & \cdots & a_0 & 0 & \cdots & \cdots & 0 \\ 0 & a_n & a_{n-1} & \cdots & \cdots & a_0 & \cdots & \cdots & 0 \\ 0 & 0 & a_n & \cdots & \cdots & \cdots & a_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_n & \cdots & \cdots & \cdots & a_0 \\ b_k & b_{k-1} & \cdots & \cdots & \cdots & b_0 & 0 & \cdots & 0 \\ 0 & b_k & b_{k-1} & \cdots & \cdots & \cdots & b_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_k & b_{k-1} & \cdots & \cdots & \cdots & b_0 \end{vmatrix}$$

称为 $p(x)$和 $q(x)$关于 x 的结式，记为 $\mathrm{Res}(p,q,x)$. 如果 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 和 $\beta_1,\beta_2,\cdots,\beta_k$ 分别是 $p(x)$和 $q(x)$的根，则有恒等式

$$\mathrm{Res}(p,q,x)=a_n^k\cdot b_k^n\prod_{i=1}^{n}\prod_{j=1}^{k}(\alpha_i-\beta_j),$$

因而也有

$$\mathrm{Res}(p,q,x)=a_n^k\prod_{i=1}^{n}q(\alpha_i).$$

可见，$p(x)$与 $q(x)$有公因式的充分必要条件是它们关于 x 的结式为 0.

升列关于多项式的结式(resultant of an ascending chain with respect to a polynomial)　结式的一种推广. 设有一升列$\{f_1,f_2,\cdots,f_s\}$和一多项式 g：

$$f_i=f_i(u_1,u_2,\cdots,u_d,x_1,x_2,\cdots,x_i)\quad(i=1,2,\cdots,s),$$
$$g=g(u_1,u_2,\cdots,u_d,x_1,x_2,\cdots,x_s).$$

令

$$R_{s-1}=\mathrm{Res}(g,f_s,x_s),$$
$$R_{s-2}=\mathrm{Res}(R_{s-1},f_{s-1},x_{s-1}),$$
$$\cdots\cdots$$
$$R_1=\mathrm{Res}(R_2,f_2,x_2),$$
$$R_0=\mathrm{Res}(R_1,f_1,x_1),$$

则称 R_0 为升列$\{f_1,f_2,\cdots,f_s\}$关于多项式 g 的结式，记为 $\mathrm{Res}(f_1,f_2,\cdots,f_s,g)$.

真升列(proper ascending chain)　一类具有较好性质的升列. 设有升列$\{f_1,f_2,\cdots,f_s\}$，其中

$$f_i=f_i(u_1,u_2,\cdots,u_d,x_1,x_2,\cdots,x_i).$$

如果对每个 i，f_i 的初式 I_i 都不是环

$$\frac{K(u_1,u_2,\cdots,u_d)[x_1,x_2,\cdots,x_{i-1}]}{(f_1,f_2,\cdots,f_{i-1})}$$

中的零元素或零因子，则称$\{f_1,f_2,\cdots,f_s\}$为真升列. 更具体地，设 f_i 关于 x_i 为 m_i 次，则方程组

$$\{f_i=0,\ i=1,2,\cdots,s\}$$

一般有 $m_1m_2\cdots m_s$ 组解

$$\{x_1^{(i_1)}(u),x_2^{(i_1,i_2)}(u),\cdots,x_s^{(i_1,i_2,\cdots,i_s)}(u)\},$$

其中 $u=(u_1,u_2,\cdots,u_d)$，$i_k=1,2,\cdots,m_k$. 如果 I_1 不是零多项式，并且对 $i=2,3,\cdots,s$，方程组

$$\{f_j=0,\ j=1,2,\cdots,i-1\}$$

的 $m_1m_2\cdots m_{i-1}$组解中每一组都不使

$$I_i=I_i(u,x_1,x_2,\cdots,x_{i-1})$$

恒为 0，则$\{f_1,f_2,\cdots,f_s\}$为真升列. 升列$\{f_1,f_2,\cdots,f_s\}$为真升列的充分必要条件是 $I_i\neq0$ 且

$$\mathrm{Res}(f_1,f_2,\cdots,f_{i-1},I_i)\neq0\quad(i=2,3,\cdots,s).$$

真升列的重要性在于，它把一些表面上是升列但性质异于一般升列的多项式组区别开来. 例如，按升列定义，多项式组

$$f_1=x_1^2-2ux+u^2,$$
$$f_2=(x_1-u)x_2+v$$

是升列，但若设 $f_1=0$，则有 $f_2=v$，f_2 的主元实质上不是 x_2 了，原因在于$\{f_1,f_2\}$不是真升列. 例如，要检验在假设条件

$$h_1=x_1^2-2ux+u^2=0,$$
$$h_2=(x_1-u)(x_2-x_1)=0$$

之下是否有

$$g=x_2-x_1=0.$$

用吴方法得出断言：结论一般成立，非退化条件是 $x_1-u\neq0$. 实际上，由于非退化条件 $x_1-u\neq0$ 与假设 $h_1=0$ 矛盾，此命题结论一般不成立. 故严格说来，吴方法要用于真升列才合理. 不过，在大量常见的几何命题中，所遇到的升列多为真升列，不是真升列的情形极为少见(参见“代数方程组相关性”).

代数方程组相关性(dependency between algebraic equations)　一多项式组与另一多项式关系的一种描述. 设有一多项式组 PS 和一多项式 g，欲问

PS 的零点集或零点集的某些子集是否是 g 的零点集的子集，这即是方程组与方程 g 的相关性问题. 在里特-吴整序原理的基础上，相关性问题可以有更清晰的表述：设已给升列

$$\mathrm{AS}=\{f_i(u_1,u_2,\cdots,u_d,x_1,x_2,\cdots,x_i),\ i=1,2,\cdots,s\}$$

和多项式 $g(u_1,u_2,\cdots,u_d,x_1,x_2,\cdots,x_i)$，并设 f_i 关于主元 x_i 为 m_i 次，则 AS 的零点集可以分为 $m_1m_2\cdots m_s$ 个支，每一支可用以 $u_1,u_2,\cdots,u_d$ 为变元的一组代数函数

$$\{x_1^{(i_1)},x_2^{(i_1,i_2)},\cdots,x_s^{(i_1,i_2,\cdots,i_s)}\}$$

表示，欲问这 $m_1m_2\cdots m_s$ 个零点集分支中，有多少个分支是 g 的零点集的子集，这是代数方程组相关性问题的一种明确而自然的提法，也是定理机器证明的一个基本理论问题. 张景中、杨路等人提出的 λ 结式法，完全地解决了这一问题：设 AS 是真升列，引进独立变元 λ，作 AS 关于 $g+\lambda$ 的结式

$$\mathrm{Res}(f_1,f_2,\cdots,f_s,g+\lambda),$$

它是关于 λ 与 $u_1,u_2,\cdots,u_d$ 的多项式. 设此多项式关于 λ 的最低次数为 k，则在 AS 的零点集的 $m_1m_2\cdots m_s$ 个分支中，恰有 k 个是 g 的零点集的子集. 这一方法可以在微机上实现，用以处理难度较大的几何定理的机器证明.

一般点(generic point) 具有特定性质的点. 设 CS 是一不可约特征列，它刻画了一个一般点 ζ，该点具有下列性质：对任意多项式 F，F 对 CS 的余式为零，当且仅当$F(\zeta)=0$.

例证法(method to proving by single-instance) 亦称单点例证法. 一种证明方法. 是通过检验一个数值例子而判定命题真假的方法. 与用一般点检验多项式对升列的余式是否为 0 的思想有类似之处，但又有实质的不同. 这一方法是洪加威提出的. 但洪加威的理论较繁且仅能用于一类构造性几何命题的判定. 下面说明基于 λ 结式的方法(参见“代数方程组相关性”).

设几何命题已转化为代数命题：要判断升列 $f_1,f_2,\cdots,f_s$ 的零点集各分支中有多少个分支包含于 g 的零点集中，若所有分支都包含于 g 的零点集中，则几何命题一般成立. 否则，命题一般不成立或在排除若干情形后成立. 而 g 的零点集包含的升列零点集分支数目，即多项式

$$P(u,\lambda)=\mathrm{Res}(f_1,f_2,\cdots,f_s,g+\lambda)$$

关于 λ 的最低次数. 由于 $P(u,\lambda)$ 中诸 λ^k 项的系数都是独立变元 $u=(u_1,u_2,\cdots,u_d)$ 的多项式，故问题归结为一些以 $u_1,u_2,\cdots,u_d$ 为变元的多项式是否恒为 0 的问题. 取定 $u_1,u_2,\cdots,u_d$ 的一组数值，求出升列 $f_1,f_2,\cdots,f_s$ 对应于这组参数的零点代入 g，即可根据 λ 结式理论检验 $P(u,\lambda)$ 中各 λ^k 项的系数对应于这组参数的值是否为 0. 根据下述定理，对某些特殊的参数值，这种检验可用于判定被检验的多项式是否恒等于 0.

定理：设 $P(u_1,u_2,\cdots,u_m)$ 是 $u_1,u_2,\cdots,u_m$ 的多项式，它关于 u_k 的次数不大于 $n_k(1\leqslant k\leqslant m)$. 又设它的标准展开式中非零系数的绝对值不大于 L，不小于 $S>0$，如果变元的一组值 $\hat{u}_1,\hat{u}_2,\cdots,\hat{u}_m$ 满足

$$\begin{cases}|\hat{u}_1|=q_1\geqslant\dfrac{L}{s}+1,\\ |\hat{u}_k|=q_k\geqslant q_{k-1}^{n_{k-1}+1}+1\ \ (k=2,3,\cdots,m),\end{cases}$$

则有 $|f(\hat{u}_1,\hat{u}_2,\cdots,\hat{u}_m)|\geqslant S>0$.

这一定理表明，如果 $P(u_1,u_2,\cdots,u_m)$不恒为 0，总可以适当取一组 $u_1,u_2,\cdots,u_m$ 的值代入，使其绝对值大于某常数. 反之，若其绝对值小于某常数，即可断言它恒等于 0. 这一定理同时解决了计算机本身不能判断结果是 0 还是一个很小的正数的问题. 由于所构思的数值例子涉及很多位的数的计算，目前尚不能用单点例证法证明难度稍大的平面几何定理.

数值并行法(parallel numerical method of mechanical theorem proving or the method to proving by multi-instance) 亦称多点例证法. 一种证明方法. 是通过检验多个数值例子而判定命题真假的方法. 这一方法是张景中、杨路提出并实现的，用它能在内存仅 4B 的袖珍计算机上证明非平凡的几何定理及发明新定理. 其理论根据是：

1. 代数方程组相关性的判定问题，总可以通过 λ 结式方法化为一个或一组仅含独立变元的多项式是否恒等于 0 的判定问题.

2. 多项式是否恒等于 0 的判定问题，可以用若干个数值例子的检验来实现. 即应用下述定理：设 $f(x_1,x_2,\cdots,x_m)$是 $x_1,x_2,\cdots,x_m$ 的多项式，它关于 x_k 的次数不大于 n_k. 对应于 $k=1,2,\cdots,m$，取数组 $a_{k,l}(l=0,1,2,\cdots,n_k)$，使得 $l_1\neq l_2$ 时 $a_{k,l_1}\neq a_{k,l_2}$，如果对任一组$\{l_1,l_2,\cdots,l_m,\ 0\leqslant l_k\leqslant n_k\}$，都有

$$f(a_{1,l_1},a_{2,l_2},\cdots,a_{m,l_m})=0,$$

则 $f(x_1,x_2,\cdots,x_m)$是恒为 0 的多项式.

称定理中所有变元组$(a_{1,l_1},a_{2,l_2},\cdots,a_{m,l_m})$组成的集合是一个规模为

$$(n_1+1)\times(n_2+1)\times\cdots\times(n_m+1)$$

的格阵，则此定理可简单地表述为：$f(x_1,x_2,\cdots,x_m)$恒为 0 的充分必要条件是它在一个规模为

$$(n_1+1)\times(n_2+1)\times\cdots\times(n_m+1)$$

的格阵上为 0.

数值并行法的工作步骤是：

1. 化几何命题为代数命题.

2. 化代数命题为代数恒等式的检验问题. 这可

以通过整序及应用 λ 结式方法完成，也可以通过其他途径. 但不论用什么方法，只需估计所化成的恒等式检验问题中待检验多项式的各变元的次数的上界，而不必真地去运算.

3. 根据前一步所估计的各变元次数的上界，构造独立变元组的格阵，并在此格阵上检验命题，如命题对格阵中每组变元为真，则一般真.

在把代数命题转化为代数恒等式检验问题过程中，曾作为分母的多项式应不为 0，这就是非退化条件.

非退化条件(nondegenerate condition)　使几何命题不失一般性的条件. 吴文俊最先明确指出：几何中一个有意义的定理，往往要附加若干限制条件，如平行四边形 $ABCD$ 的对角线 AC 与 BD 互相平分，其限制条件是 $ABCD$ 必须是非退化的平行四边形，即 A,B,C,D 四点不能在同一直线上. 这样的限制条件就称为几何定理的非退化条件.

吴方法的精美处之一就是，整序过程中可以确定出一类非退化条件. 具体地，从假设条件的多项式组经整序得到升列 $f_1,f_2,\cdots,f_s$ 后，记 f_j 的初式为 I_j，则吴氏非退化条件为 $I_1I_2\cdots I_s\neq 0$. 当结论多项式 g 关于升列 $f_1,f_2,\cdots,f_s$ 的余式为 0 时，升列 $f_1,f_2,\cdots,f_s$ 的任一零点，只要满足非退化条件，就一定是 g 的零点. 若不满足非退化条件，则可能不是 g 的零点. 吴氏非退化条件另一应用是：若升列是由一多项式组 PS 整序而得到，则升列的零点满足非退化条件时，必为 PS 的零点.

使用格若勃基方法证明定理时，也会产生非退化条件——要求基的表达式中分母不为 0. 从同一多项式出发，所得到的吴氏非退化条件与格若勃基非退化条件不一定相同.

弱非退化条件(weak nondegenerate condition)　吴氏非退化条件的改进. 由张景中、杨路等人提出. 吴氏非退化条件要求升列中多项式初式不为 0，即多项式关于主变元的最高次项系数不为 0. 弱非退化条件则只要求主变元各项系数不同时为 0. 例如，设

$$f_1=x^2-u,$$
$$f_2=(x+2u)y^2-(3x+2u)y+(u+1),$$
$$g=u(2x+1)y^2-(3+2x)uy+x(u+1).$$

欲判断在条件 $f_1=f_2=0$ 之下是否有 $g=0$，可按吴方法求 g 关于升列 $\{f_1,f_2\}$ 的余式，结果余式为 0. 于是得到结论：在非退化条件 $x+2u\neq 0$ 满足时，由 $f_1=f_2=0$ 可得 $g=0$. 但由弱非退化条件，此结论可改进为：若 $(x+2u)$，$(3x+2u)$，$(u+1)$ 不同时为 0，则由 $f_1=f_2=0$ 可得 $g=0$. 但又因 $(x+2u)$，$(3x+2u)$，$(u+1)$ 显然不能同时为 0，故得出结论"由 $f_1=f_2=0$ 可得 $g=0$".

几何数系(number system about a geometry)　具有某种性质的代数系统. 不同的几何体系由于满足的公理系统不同，因而所决定的代数系统具有的运算性质也不尽相同. 这些不同的代数系统称为相应的几何数系. 几何数系不同时，几何问题的代数化方法可能不同，其中定理机器证明的方法也会有所不同.

可读证明(readable proof)　机器证明的基本概念. 指能够不花太多的时间和精力就可以被人理解和检验其正确性的证明. 用传统方法给出的几何定理的证明是可读的. 而用机器证明的代数方法，如吴方法、GB 法或数值并行法来证明定理，由于过程中计算量一般较大，人们难以看明白并加以检验，所产生的证明不是可读的. 能不能用机械的方法产生可读证明？自 1960 年以来，欧美一些学者开始研究这一重要问题，提出了各种设想. 中国的吴方法成功后，这个问题引起了人们的更大关注. 1992 年 5 月，张景中等人给出了世界上第一个能自动产生几何定理可读证明的算法和通用程序. 在微机上成功地产生了近千条非平凡的几何定理的可读证明，其中许多证明简短而优美，这是机器证明的又一突破. 机器生成可读证明的研究在几年间得到很大的发展.

系统面积方法(systematic area method)　利用面积关系解决几何问题的一套具有通用性的方法. 其主要工具是两个命题：

1. 共边定理：若两直线 AB，PQ 交于 M，则 $\triangle PAB$ 与 $\triangle QAB$ 的面积比等于两线段 PM 与 QM 的比.

2. 共角定理：若 $\angle ABC$ 和 $\angle XYZ$ 相等或互补，则 $\triangle ABC$ 与 $\triangle XYZ$ 的面积比等于线段乘积 $AB\times BC$ 与 $XY\times YZ$ 的比.

这套方法，特别是共边定理的运用，为机器产生可读证明的第一个有效的算法提供了几何处理手段.

勾股差(pythagoras difference)　利用面积关系解决几何问题的基本工具之一，也是机器生成可读证明的第一个成功算法中要用的基本几何量之一. 对平面上任意 $\triangle ABC$，关于 $\angle B$ 的勾股差定义为 $P_{ABC}=AB^2+BC^2-AC^2$(即 $P_{ABC}=2BC\cdot BA\cdot\cos\angle ABC$)，而四边形 $ABCD$ 的勾股差则定义为

$$P_{ABCD}=AB^2-BC^2+CD^2-DA^2.$$

它们之间的关系为

$$P_{ABCD}=P_{ABD}-P_{CBD},\quad P_{ABBC}=P_{ABC}.$$

四边形勾股差 P_{ABCD} 和带号面积 S_{ABCD} 有许多类似的性质，事实上，它是向量 $\overrightarrow{CA}$ 与 $\overrightarrow{BD}$ 的内积的两倍. 在机器证明中，勾股差的最基本应用在于：

1. 直线 AC 与 BD 垂直的充分必要条件是

$$P_{ABCD}=0.$$

2. 如果 Q 到直线 AB 的垂足为 D，则

$$\frac{AD}{DB}=\frac{P_{QAB}}{P_{QBA}}.$$

用这个等式可消去垂足 D.

3. 若 AB 的垂线过点 P 而与 UV 交于 X，则

$$\frac{UX}{XV}=\frac{P_{UAPB}}{P_{PAVB}}.$$

用这个等式可消去交点 X. 四边形勾股差在立体几何机器证明中也有重要用处，这主要是因为(异面或共面)两直线 AC，BD 所成锐角的余弦值恰为

$$\left|\frac{P_{ABCD}}{2AC\cdot BD}\right|.$$

消点原理(principle to eliminate point)　几何定理可读证明机器生成算法的基本思想. 其要点是：用一些几何量之间的代数关系来表达命题的结论，再利用命题的假设条件从这些几何量中逐个地消去受条件约束的点. 每消去一个点，便得到一个与原命题等价的新命题，但新命题的假设条件中所涉及的点比原命题减少了. 当点减少到一定程度时，命题的真假便成为显然的了. 消点过程中所产生的命题系列，便构成了命题的一个证明或反证. 由于消点操作一般是直接利用几何量之间的关系实现的，不化为点的坐标运算，所以产生的证明通常是简明可读的.

几何定理可读证明的自动生成(automated production of readable proofs for the theorems in geometry)　机械地证明几何定理并生成可读证明的方法. 这方面的研究自 1960 年以来已进行了 30 多年，提出了各种设想. 第一个成功的方法是 1992 年由张景中等人在微机上实现为通用程序的. 算法设计的原理是：

1. 把命题中涉及的点依照作图时出现的顺序排列，使每个点或者是任取的自由点，或者是与已经做出的某些点有关的约束点，使命题的假设转化为作图的过程.

2. 把命题的结论用特定的几种几何量之间的一个代数关系来表达. 例如，在第一个成功的算法中，用到了三种几何量，即四边形和三角形的带号面积、四边形和三角形的勾股差，以及平行或共线的有向线段的比.

3. 从命题的结论中自后而前逐个消去在作图时出现的点，直到水落石出. 具体的消点公式依赖于该点由哪种作图产生和出现在哪种几何量中. 例如，当采用八种基本作图和三种几何量时，至多用 24 个消点公式就够了.

4. 为了使产生的证明更为简明优美，算法中还使用了几何数据库的技巧：从命题的假设条件中，运用一些基本的几何知识，推出许多与命题图形有关的几何性质，构成一个临时的数据库备用(如基本的共线点、平行线、垂直线、共圆点等). 利用这个数据库常常可以简化消点公式，以下用两个例子来说明. 例子中记号 S_{ABC} 表示△ABC 的带号面积，S_{ABCD} 表示四边形 $ABCD$ 的带号面积，即

$$S_{ABCD}=S_{ABC}+S_{ACD}=S_{ABD}+S_{BCD}.$$

而记号 P_{ABC} 表示△ABC 关于∠B 的勾股差，即

$$P_{ABC}=AB^2+BC^2-AC^2$$
$$=2BC\cdot BA\cdot\cos\angle ABC.$$

记号 P_{ABCD} 则表示四边形 $ABCD$ 的勾股差，即

$$P_{ABCD}=AB^2-BC^2+CD^2-DA^2.$$

此法效率很高，已用于证明近千条几何定理，并被推广到非欧几何，发现一批新定理.

全角(fully angle or total angle)　描述平面上两直线相对位置的一种几何量. 同一平面上的两条直线 l_1，l_2 所构成的有序对称为 l_2 的全角，记为 $\langle l_1\ l_2\rangle$. 若 l_1，l_2 分别为 AB，CD，则 $\langle l_1\ l_2\rangle$ 也可记为 $\langle AB\ CD\rangle$，$\langle AB\ l_2\rangle$ 或 $\langle l_1\ CD\rangle$. 若 $l_1\perp l_2$，则称 $\langle l_1\ l_2\rangle$ 为直角，记为 $\langle l_1\ l_2\rangle=\langle 1\rangle$；若 $l_1/\!/l_2$ 或两者重合，称 $\langle l_1\ l_2\rangle$ 为平角，记为 $\langle l_1\ l_2\rangle=\langle 0\rangle$.

全角在几何定理机器证明中有重要应用. 吴文俊首先在他的机器证明方法中使用全角概念，但他引入的全角定义中要求两直线相交，与此处略有不同，中学解析几何教材中直线 l_1 到 l_2 的角，实质上是全角.

直线 l_1 依反时针方向旋转到首次与 l_2 平行的位置时所旋过的角的值，称为全角 $\langle l_1\ l_2\rangle$ 的值. 取直角为基本单位时，n 的全角值为 $\langle n/90°\rangle$. 如 $\langle 1/2\rangle=\langle 45°\rangle$，$\langle 1/3\rangle=\langle 30°\rangle$. 全角可在 $[0,2)$ 中或 $(-1,1]$ 中取值，值相等(模 2)的全角称为相等的全角. 全角之间的加减对应于它们值的模 2 加减，有交换、结合等运算律. 在机器证明中最常用的还有过渡律：

$$\langle l_1\ l_2\rangle+\langle l_2\ l_3\rangle=\langle l_1\ l_3\rangle.$$

全角是有方向性的，$\langle l_1\ l_2\rangle=-\langle l_2\ l_1\rangle$，即

$$\langle l_1\ l_2\rangle+\langle l_2\ l_1\rangle=\langle l_1\ l_1\rangle=\langle 0\rangle,$$

但所涉及的直线是无向的，即 $\langle AB\ l\rangle=\langle BA\ l\rangle$.

全角的几何性质比传统角简明. 例如，直线 AB，CD 被第三条直线所截，交点分别为 P，Q，则 $AB/\!/CD$ 的充分必要条件是 $\langle AB\ PQ\rangle=\langle CD\ PQ\rangle$. 这避免了有关三线八角的序关系描述. 又如，四点 A，B，C，D 共圆或共线的充分必要条件是 $\langle AB\ BC\rangle=\langle AD\ DC\rangle$，避免了四点位置情形的讨论.

在机器证明中使用全角，常可产生优美而简捷的证明. 1994 年，一份研究报告中用全角法给出 106 个非平凡几何定理的机器证明. 下面的西姆森定理的全角证明是计算机生成的：已知 D 在△ABC 的外接圆上，自 D 向 BC，CA，AB 分别引垂线得垂足

E,F,G,则 E,F,G 三点在一直线上.(机器输入略)

机器生成之证明为

$\langle GF\ GE\rangle=\langle GF\ GD\rangle+\langle GD\ GE\rangle$(过渡)

$=\langle AF\ AD\rangle+\langle BD\ BE\rangle$

(因 G,F,A,D 共圆,G,D,B,E 共圆)

$=\langle AF\ AD\rangle+\langle BD\ BC\rangle$(因 B,C,E 共线)

$=\langle AC\ AD\rangle+\langle BD\ BC\rangle$(因 A,F,C 共线)

$=\langle BC\ BD\rangle+\langle BD\ BC\rangle$

(因 A,B,C,D 共圆)

$=\langle 0\rangle$.

这证明了 G,F,E 共线.在传统证明中,要根据图中点的位置,说明 G,E 在 AC 同侧或异侧才能推出所要的结论.

柱形代数剖分(cylindrical algebraic decomposition) 简称 CAD.一种用于证明不等式型定理的算法.它是科林斯(Collins,G. E.)于 1975 年提出的.设有一命题,假设其由多项式的不等式和等式构成:

$$\begin{cases}f_1(x_1,x_2,\cdots,x_n)\geqslant 0,\cdots,f_s(x_1,x_2,\cdots,x_n)\geqslant 0,\\ g_1(x_1,x_2,\cdots,x_n)>0,\cdots,g_l(x_1,x_2,\cdots,x_n)>0,\\ h_1(x_1,x_2,\cdots,x_n)=0,\cdots,h_m(x_1,x_2,\cdots,x_n)=0.\end{cases}$$

而其结论为 $\varphi\geqslant 0$ 或 $\varphi>0$ 或 $\varphi=0$.对此命题应用柱形代数剖分的大致步骤是:

1. 将 $f_i=0,g_j=0,h_k=i,\varphi=0$ 等($i=1,2,\cdots,s$,$j=1,2,\cdots,l,k=1,2,\cdots,m$)视为 n 维空间中的曲面.这些曲面将空间剖分为有限个不同维数的“胞腔”(cell),使得 f_i,g_j,h_k,φ 等每个多项式的每个胞腔中保持不变号,即恒为正,或恒为负,或恒为零.

2. 在每一个胞腔中至少选定一个测试点,设全体测试点的集为

$$S=\{p_\nu=(x_1^{(\nu)},x_2^{(\nu)},\cdots,x_n^{(\nu)}),\nu=1,2,\cdots,\mu\}.$$

3. 在 S 中挑选出所有能满足假设条件的测试点,它们组成一个子集 S^*.

4. 如果 S^* 中每一个点都能使命题的结论成立,则命题为真;否则命题为假.

这个算法相对于过去有很大的进步,但其计算复杂度仍然偏高,故用此法能在计算机上实际判定的仍局限于一些比较简单的不等式.有人指出,当空间的维数 $n>3$ 时,这样的剖分在目前的软硬件条件下一般是难于实现的.

证明不等式的降维算法(dimension-decreasing algorithm for inequality proving) 通过控制维数以提高不等式机器证明效率的算法.它是杨路于 1998 年提出的.关于几何定理机器证明的吴文俊方法的成功,在世界范围内推动了计算机推理研究的加速发展.但在过去的 20 年中,等式型和不等式型定理机器证明两类算法的效率之间出现了极大的差距,对几何定理而言,差距尤为明显.降维算法和相应的程序是为缩小这一差距所作的努力.不等式机器证明的困难在于它依赖于实代数的自动推理,其复杂度随维数增加而急速升高.当命题的假设中含有若干代数方程时,一个自然的想法是消去一些变元以降低维数.这样做所遭遇的新的困难是,必须有效处理带参数的根式或其他非有理式的代数函数.降维算法正是一个既能将维数保持在最低限度,又能有效地将非多项式代数函数有理化的方法.所用基本概念如下:

1. 给定一个以 x,y,z 等为变量的代数不等式 F,其左边或右边都可能是非有理式的代数函数.如果$L(T)$是 T 的多项式,而它所有系数都是 x,y,z 等的多项式,且 F 的左边是 $L(T)$的一个根,称 $L(T)$是 F 的一个左多项式.类似地可以定义 F 的右多项式 $R(T)$.

2. 以 $P(x,y,z,\cdots)$表示 $L(T)$与 $R(T)$关于 T 的结式,并将 $P(x,y,z,\cdots)=0$ 定义的曲面称为不等式 F 的临界曲面.

降维算法概略介绍如下:

1. 确定所要验证的不等式的临界曲面.

2. 临界曲面将参数空间剖分为有限个部分,在每一部分任意选定至少一个测试点.

3. 只需对这有限个测试点验证该不等式.

如果对这些特殊的参数值命题为真,则命题普遍真.根据以上算法用 MAPLE 语言编制成的通用程序 BOTTEMA 已在 PC 机上成功地验证了 1000 多个不等式,其中近半数是近期提出未加证实的猜想.对名著《几何不等式》中 120 个基本不等式的验证总共费时仅 20 多秒.运用该软件已能成批验证非平凡的不等式.就两个自由度的情形,其效率与等式型几何定理证明软件相比已在伯仲之间.这一情况使得若干多项式型的全局优化问题可以转化为对有限多个不等式的验证.

多项式完全判别系统(complete discrimination system for polynomials) 一组由多项式的系数构成的显式表达式.利用它可以完全确定此多项式根的分类,即多项式的实根个数、虚根个数以及它们的重根数.对于涉及不等式的机器证明有重要意义.二次实系数多项式 $ax^2+bx=c$ 的完全判别系统就是它的判别式 b^2-4ac.但对次数较高的多项式,实系数四次多项式的完全判别系统是 20 世纪 60 年代才完成的.实系数五次多项式的完全判别系统直到 1995 年还未找到.1996 年,杨路等人建立了实系数多项式的完全判别系统以及根的详尽分类的显式判准,圆满解决了这个实代数的基本问题.1999 年,张景中等人建立了复系数多项式的完全判别系统,进一步解决了复系数多项式根的完全分类问题,而且

把所得算法用 MAPLE 编成一个通用程序，使多项式的完全判别系统及其完全根分类由计算机自动生成，不需要人工的干预.

复系数多项式根的分类情况由下述定理给出：设 $f(x)=f_1(x)+I^* f_2(x)$，$f_i(x)$是实系数多项式$(i=1,2)$. 令

$$p(x)=\gcd(f_1(x),f_2(x)),$$

$$g(x)=\frac{f(x)}{p(x)},$$

则 $p(x)$的所有根恰由$f(x)$的所有实根和所有共轭虚根组成，而 $g(x)$的所有根恰由 $f(x)$的所有非共轭虚根组成. 设 $f(x)=a_0x^n+a_1x^{n-1}+\cdots+a_n$. $f(x)$与 $f'(x)$的西尔维斯特矩阵

$$\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_n & 0 & & 0 \\ 0 & na_0 & (n-1)a_1 & \cdots & a_{n-1} & & & 0 \\ & a_0 & a_1 & \cdots & a_{n-1} & a_n & & 0 \\ & 0 & na_0 & \cdots & 2a_{n-2} & a_{n-1} & & \\ & & & \cdots & \cdots & & & \\ & & & \cdots & \cdots & & & \\ & & & & a_0 & a_1 & \cdots & a_n \\ & & & & 0 & na_0 & \cdots & a_{n-1} \end{bmatrix}$$

称为$f(x)$的判别矩阵，并记为 Discr(f). 用 D_k 表示由 Discr(f)的前 $2k$ 行、前 $2k$ 列所形成的子矩阵的行列式，$k=1,2,\cdots,n$. n 元组$[D_1,D_2,\cdots,D_n]$称为$f(x)$的判别式序列. 设$[D_1,D_2,\cdots,D_n]$是 $p(x)$的判别式序列，记

$$[s_1,s_2,\cdots,s_n]$$
$$=[\mathrm{sign}(D_1),\mathrm{sign}(D_2),\cdots,\mathrm{sign}(D_n)],$$

则按如下的方法构造一个新表$[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n]$，称它为 $p(x)$所对应的符号修订表. 如果$[s_i,s_{i+1},\cdots,s_{i+j}]$是符号表$[s_1,s_2,\cdots,s_n]$中的一段，满足

$$s_i\neq0,\ s_{i+1}=s_{i+2}=\cdots=s_{i+j-1}=0,\ s_{i+j}\neq0,$$

则对 $\gamma=1,2,\cdots,j-1$，令

$$\varepsilon_{i+\gamma}=(-1)^{\left[\frac{\gamma+1}{2}\right]}\cdot s_i;$$

而符号表$[s_1,s_2,\cdots,s_n]$的其他元素不变，即令 $\varepsilon_k=s_k$. 设$[d_1,d_2,\cdots,d_t]$为$g(x)$的判别式序列，如果它不为零的元素的最大下标为 k，则称表$[1,1,\cdots,1,0,0,\cdots,0]$(连续 k 个 1，紧接着连续 $t-k$ 个 0)为 $g(x)$所对应的符号修订表.

如果单求互异根个数，那么下面的命题就足够了：给定实系数多项式 $p(x)$，如果它所对应的符号修订表之变号数及非零元素个数分别为 ν 和 s，则 $p(x)$有 ν 对互异共轭虚根，有 $s-2\nu$ 个互异实根. 设

$$g(x)=g_1(x)+I^* g_2(x),$$
$$\gcd(g_1(x),g_2(x))=1,$$

如果$g(x)$所对应的符号修订表中非零元素的个数为 γ，则$g(x)$有 γ 个互异根，且全是非共轭虚根.

如果不仅仅求 $f(x)$互异根的个数，还要求每个根的重数，那么还需考虑 $\gcd(f(x),f'(x))$的根分类情况. 用$\triangle(f)$表示$\gcd(f(x),f'(x))$，并记

$$\triangle^0(f)=f(x),\ \triangle^j(f)=\triangle(\triangle^{j-1}(f)),$$

称

$$\{\triangle^0(f),\triangle^1(f),\triangle^2(f),\cdots\}$$

为 $f(x)$的$\triangle$序列. 一个多项式的$\triangle$序列是有限序列. 设$p(x)$和$g(x)$的$\triangle$序列分别为 U 和 V，而用 W 表示它们的并集. W 中所有多项式的判别式序列将构成 $f(x)$的完全判别系统. 下面给出由完全判别系统确定多项式 $f(x)$根分类的方法步骤. 以形式"[[实根分类],[共轭虚根分类],[非共轭虚根分类]]"表示给定多项式的一个根分类：

1. 求出 $p(x)$和 $g(x)$所对应的符号修订表. 对 $p(x)$来说，计算符号修订表中非零元素的个数以及变号数，以确定它的不同实根和共轭虚根的数目；对 $g(x)$来说，计算符号修订表中非零元素的个数，以确定它不同的非共轭虚根数目. 如果符号修订表中不含 0，则结束.

2. 如果上面的符号修订表含有 k 个 0，那么$\triangle(p)$(或$\triangle(g)$)是一个 k 次多项式. 对$\triangle(p)$(或$\triangle(g)$)，执行与第一步中对 $p(x)$(或 $g(x)$)相同的操作.

3. 以同样方式处理$\triangle^2(p)$，$\triangle^3(p)$，…(或$\triangle^2(g)$，$\triangle^3(g)$，…)，直到对某个 j，$\triangle^j(p)$(或$\triangle^j(g)$)所对应的符号修订表不含 0.

4. 对$\triangle^j(p)$(或$\triangle^j(g)$)，用第一步的方法计算各种类型根的数目以及各个根的重数；然后计算$\triangle^{j-1}(p)$(或$\triangle^{j-1}(g)$)各种类型根的数目以及各个根的重数；然后对$\triangle^{j-2}(p)$(或$\triangle^{j-2}(g)$)进行相同的操作……，直到获得 $p(x)$(或 $g(x)$)的完全根分类.

多项式的判别矩阵(discrimination matrix for polynomials)　见"多项式完全判别系统".

多项式的判别式序列(discrimination serial for polynomials)　见"多项式完全判别系统".

撰　稿　朱瑞香　刘卓军　杨　路　张永奎　张永朝
张景中　陈有祺　周玉龙　骆成熙　黄昌宁
梁松新　董士海
审　阅　石纯一　许卓群　杨芙清　李　未　沈美明
张乃孝　张景中　郑纬民　董士海

数 理 语 言 学

数理语言学(mathematical linguistics)　数学科学的边缘学科.它是用数学方法和数学思想研究语言现象的一门新兴学科,它使语言学与数学、计算机科学、控制论以及人工智能发生了密切的联系.

数理语言学产生于20世纪50年代.1955年,美国哈佛大学首先创办了数理语言学讨论班,1957年正式开设数理语言学课程.此后,许多国家开展了这方面的教学和研究.1982年,中国在北京大学开设了"语言学中的数学问题"的选修课,首次在国内讲授数理语言学.

关于语言问题的数学研究,早在数理语言学这个学科的名称正式出现之前就开始了.1847年,俄国数学家布尼亚科夫斯基(Буняковский, В. Я.)提出可以用概率论来进行语法、词源和语言历史演变的研究.1851年,英国数学家德·摩根(De Morgen, A.)把词长作为文章风格的一个特征进行统计研究.1867年,苏格兰学者坎贝尔(Campbell,L.)用统计方法来确定古希腊学者柏拉图(Plato)著作的执笔时期.1881年,德国学者迪丁贝尔格(Dittinberger, W.)进一步用统计方法把柏拉图著作的执笔时期分为前期、中期和后期三个阶段.1887年,美国学者门登豪尔(Mendenhall,T.C.)对不同时期的英国文学著作进行统计分析,特别是研究了莎士比亚(Shakespeare,W.)的作品.1894年,瑞士语言学家索绪尔(Saussure, F. De)认为,可以用数学公式来有规律地表达语言中量与量之间的关系,他又把语言学与几何系统和只有复杂项的代数相比.1898年,德国语言学家凯定(Kaeding, F. W.)编制了世界上第一部频率词典《德语频率词典》,用以改进德语速记方法.1904年,波兰语言学家博杜恩·德·库尔特内(Baudouin de Courtenay, J. N.)认为,语言学家应该掌握初等数学和高等数学,语言学将根据数学的模式更多地扩展量的概念,并发展演绎思想的方法.1913年,俄国数学家马尔可夫(Марков, А. А.)研究了普希金(Pushkin,A.)的叙事长诗《欧根·奥涅金》中俄语元音和辅音字母序列,提出了马尔可夫随机过程的数学理论.1925年,中国教育家陈鹤琴发表了第一部汉字频率统计的著作《语体文应用字汇》.1933年,美国语言学家布龙菲尔德(Bloomfield,L.)认为数学是语言所能达到的最高境界.1935年,美国语言学家齐普夫(Zipf, G. K.)提出了齐普夫定律.同年,加拿大学者贝克(Beke, E. V.)提出了词的分布率的概念,并以之作为词典选词的主要标准.1944年,英国数学家尤尔(Yule, G. U.)出版了《文学词语的统计分析》一书,大规模地使用概率和统计的方法来研究语言.1949年,法国学者米谢阿(Michea,R.)提出了建立词汇统计学的设想.1950年,美国语言学家斯瓦德士(Swadesh, M.)进行了语言年代学的研究.1951年,美国数学家仙农(Shannon,C.E.)用信息论的方法来研究书面英语中的熵和多余度.同年,美国语言学家英格维(Yingve, V.)对句法现象进行了间距分析.1954年,法国学者基罗(Guiraud, P.)根据文章中词的频率分布提出了词汇丰富度的概念.

由于通信技术的发展,需要寻求语言的最佳编码的方法,以提高信道的传输能力,因而要对语言的统计特性进行精密的研究.机器翻译、情报自动检索等技术的出现,又要求精确地描述和解释语言的结构,建立语言的数学模型,并用数学方法来研究语言的语法和语义.实践中提出的这些要求,使语言文字的数学研究更加必要,而科学技术的发展也为在语言文字研究中使用精确的数学方法提供了可能.20世纪以来迅速发展的概率论、数理统计、信息论、集合论、数理逻辑、图论、格论、模糊数学和抽象代数等数学学科,为用数学方法研究语言文字提供了有力的工具.近40年来,由于在语言研究中广泛地使用电子计算机,逐渐改变了手工查频、手工统计的方法,为语言文字的数学研究提供了现代化手段.而在传统语言学中提出的分析句法以及在结构语言学中提出的以替换和分布为手段来辨别语素、分析层次的严格的语言研究方法,也为语言文字的数学研究创造了条件.

数理语言学的研究主要包括三方面内容:

1. 代数语言学.
2. 统计语言学.
3. 应用数理语言学.

代数语言学是采用集合论、数理逻辑、算法理论、模糊数学、图论、格论等离散的、代数的方法来研究语言.统计语言学是采用概率论、数理统计和信息论等统计数学的方法来研究交际过程中语言成分使用的统计规律.而把代数语言学和统计语言学应用于机器翻译、人机对话以及情报自动检索的技巧和方法,就是应用数理语言学的研究内容.

代数语言学的目的在于建立语言的数学模型,对客观的语言现实进行抽象的代数描述和理论上的精确分析,从而把语言文字研究的某些方面改造成

数学那样的演绎系统.代数语言学中的语言模型主要有分析性模型、生成性模型和辨识性模型三种.分析性模型主要采用集合论和数理逻辑的方法对语法的基本概念进行数学描述;生成性模型着重于研究形式语法和自动机理论,阐明形式语法及其与自动机的关系;辨识性模型则研究句法类型演算的方法.此外,还有把句法与语义结合起来的蒙塔鸠语法,20世纪80年代以来,首先从美国学者伍兹(Woods, W.)的扩充转换网络开始,在美国计算语言学家布列斯南(Bresnan, J.)的关于面向词汇的非转换语言学思想的激励之下,美国计算语言学家卡普兰(Kaplan,R.)和布列斯南一起于1983年提出了词汇功能语法;美国计算语言学家凯依(Kay, M.)于1983年提出了合一语法,又于1985年提出了功能合一语法,这些语法都采用复杂特征集来描述语言的结构,并且用合一运算来改进集合论中的求并运算.这些语言模型和语言理论的研究,冲破了传统语言学归纳方法的束缚,给语言学的研究带来了变革性的重大影响,而且语言现象经过模型化和数学方法的描述之后,更便于计算机的处理,为自然语言的自动处理提供了手段.

统计语言学的目的在于建立语言的统计模型.采用统计方法来研究语言的音素和音位的,有语音统计学;研究亲属语言分化年代的,有语言年代学;研究文章风格的,有风格统计学;此外,还统计语言文字的字频、词频和统计分布.

包括机器翻译、自然语言理解、情报自动检索在内的应用数理语言学,其主要目的是研究语言自动分析和自动生成的方法,这方面的研究目前已经取得显著的成绩.

中国早在20世纪20年代就进行过汉语文字的数学研究,当时采用手工查频方法.后来,台湾交通大学用了2000多人日,根据200余万字的资料用手工进行过汉字的频率统计.北京新华印刷厂动员了19个单位,用手工方法花了两年多的时间,对2100多万汉字的资料进行汉字查频工作.20世纪70年代以来,中国开始利用电子计算机进行汉语汉字的统计研究,除了统计汉字的字频和单词的词频之外,还统计了汉字的部件和笔画的频率,计算出汉字的熵和极限熵,并建立了汉语语料库,进行逐字、逐词索引和文本自动标注,统计汉语句型的频率,取得了可观的成绩.这些研究成果为外汉和汉外机器翻译系统、汉语理解系统、汉语自动情报检索系统等自然语言处理系统的开发和研制,提供了有用的信息.

代数语言学

代数语言学(algebraic linguistics)　数理语言学的一个分支学科.狭义的代数语言学把语言看成有限生成元自由幺半群的子集进行研究,是半群论的一个独特分支.广义的代数语言学指一切使用离散数学工具研究语言现象的学科,主要包括以下几个方面:

1. 建立语言的数学模型.
2. 研究形式语言及其与自动机的关系.
3. 建立自然语言自动处理的各种方法和理论.
4. 研究语音、语法、词汇、语义的模糊现象.
5. 研究语言的句法结构与语义解释之间的逻辑关系.

随着科学的发展,代数语言学正在不断开拓新的研究领域.

形式文法(formal grammar)　对于语言结构的形式描述.美国语言学家乔姆斯基(Chomsky, N.)创立的用于描述语言的文法.一部形式文法是一个四元组,即

$$G=(V_N,V_T,S,P),$$

其中,V_N 是非终极符号集,V_T 是终极符号集,V_N 与 V_T 不相交;S 是 V_N 中的一个特定元素,称初始符号;P 是一些重写规则的集合,而重写规则具有形式 $\varphi\to\psi$,这里 φ 和 ψ 都是符号串.

一个符号串 α 中如果含有一段与 φ 相同的子串 $\alpha\to\alpha_1\varphi\alpha_2$,把 φ 用 ψ 替换,得到 $\alpha'=\alpha_1\psi\alpha_2$,$\alpha'$ 称为由 α 用规则 $\varphi\to\psi$ 重写得到的符号串.若 α' 可以由 α 用 P 中的某一规则重写得到,就称 α 直接派生 α'.若 α_1 直接派生 $\alpha_2,\cdots,\alpha_{n-1}$ 直接派生 α_n,则称 α_1 派生 α_n.单个初始符号构成的符号串称为初始集合构成的一个语言,即该文法 G 描述的语言,通常记为 $L(G)$.

语言推导史(derivational history of language)　句子推导的历史.它是语言中利用形式文法推出句子的过程.例如,在英语中有如下的文法:

$$G=(V_N,V_T,S,P),$$
$$V_N=\{NP,VP,T,N,V\},$$
$$V_T=\{\text{the, man, ball, saw}\},$$
$$S=S,$$
$$P=\begin{cases} S\to NP^\frown VP & (1)\\ NP\to T^\frown N & (2)\\ VP\to V^\frown NP & (3)\\ T\to \text{the} & (4)\\ N\to \text{man, ball} & (5)\\ V\to \text{saw} & (6)\end{cases}$$

这里,初始符号 S 表示句子,NP 表示名词短语,VP 表示动词短语,T 表示定冠词,N 表示名词,V 表示动词,"⌢"表示符号串的并置.这样,"the man saw the ball"这个英语句子的推导史可以写成如下形

式：

$$S$$

$$NP^\frown VP \quad (1)$$

$$T^\frown N^\frown VP \quad (2)$$

$$T^\frown N^\frown V^\frown NP \quad (3)$$

$$\text{the } N^\frown V^\frown NP \quad (4)$$

$$\text{the man } V^\frown NP \quad (5)$$

$$\text{the man saw } NP \quad (6)$$

$$\text{the man saw } T^\frown N \quad (2)$$

$$\text{the man saw the } N \quad (4)$$

$$\text{the man saw the ball} \quad (5)$$

上下文有关文法(context-sensitive grammar) 形式文法的一种类型. 在文法 $G=(V_N,V_T,S,P)$ 中,如果其重写规则都具有形式

$$\varphi_1 A\varphi_2 \to \varphi_1\omega\varphi_2,$$

其中 A 是非终极符号,$\varphi_1,\varphi_2,\omega$ 是任意的符号串. 这样的重写规则可看成在上下文(φ_1,φ_2)中给出$A\to\omega$,那么G就称为上下文有关文法,又称 1 型文法.

上下文无关文法(context-free grammar) 形式文法的一种类型. 在文法 $G=(V_N,V_T,S,P)$ 中,如果规则的形式为 $A\to\omega$,其中 A 是非终极符号,ω 是任意符号串,那么 G 就称为上下文无关文法,又称 2 型文法.

有限状态文法(finite state grammar) 形式文法的一种类型. 在文法 $G=(V_N,V_T,S,P)$ 中,如果重写规则的形式为 $A\to aQ$ 或 $A\to a$,其中 A 和 Q 是非终极符号,a 是终极符号,那么就把这种文法称为有限状态文法,又称 3 型文法. 如果在重写规则 $A\to aQ$中,把 A 和 Q 看成不同的状态,那么当从状态 A 到状态 Q 时,就可生成一个终极符号 a,这样便可把有限状态文法想象为一种生成装置,这种装置每次只能生成一个终极符号,而每一个非终极符号都与一个特定的状态相联系,由此可以建立有限状态文法与有穷状态自动机的关系.

0 型文法(type 0 grammar) 形式文法的一种类型. 在形式文法分类中泛指任何文法. 1 型文法是 0 型文法的子类,2 型文法是 1 型文法的子类,3 型文法是 2 型文法的子类.

推导树(derivational tree) 描述句子结构的树形图. 设 $G=(V_N,V_T,S,P)$是上下文无关文法,如果有某个成分结构树满足下列条件,它就是上下文无关文法的推导树:

1. 每一个结点有一个标记,这个标记就是 $V=V_N\cup V_T$ 中的符号.

2. 根的标记是 S.

3. 如果结点 n 至少有一个异于其本身的后裔,并有标记 A,那么 A 必定是非终极符号集 V_N 中的符号.

4. 如果结点 $n_1,n_2,\cdots,n_k$ 是结点 n 的直接后裔,从左向右排列,其标记分别为 $A_1,A_2,\cdots,A_k$,那么 $A\to A_1A_2\cdots A_k$ 必定是 P 中的重写规则.

成分结构树(constituent structure tree) 数理语言学的基本概念之一. 五元组 $T=(N,Q,D,P,L)$称为成分结构树,其中:N 是结点的有限集;Q 是标记的有限集;D 是在 $N\times N$ 上的弱偏序,即支配关系;P 是在 $N\times N$ 上的严偏序,即前后关系;L 是 N 到 Q 的一个函数,即标记函数. 该五元组满足下列条件:

1. 单根条件. 只有一个结点是支配每一个结点的,这个结点就是根. 这个条件可以写为

$$(\exists x\in N)\ (\forall y\in N)(x,y)\in D.$$

2. 互斥条件. 对于任何的两个结点 x 与 y,x 与 y 处于前后关系 $P(x,y)$中,即或者$(x,y)\in P$,或者$(y,x)\in P$;当且仅当 x 与 y 不处于支配关系中,即$(x,y)\notin D$ 且$(y,x)\notin D$. 这个条件可以写为

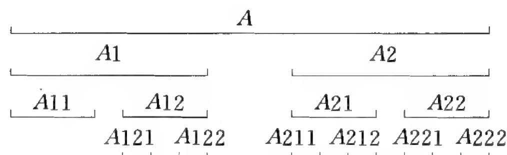
$$\{[(\forall x,y\in N)(x,y)\in P\vee(y,x)\in P] \Leftrightarrow[(x,y)\notin D\wedge(y,x)\notin D]\}.$$

3. 非交条件. 对于任何的结点 x 与 y,如果 x 前于 y,则由 x 支配的所有结点前于由 y 支配的所有结点. 这个条件可以写为

$$(\forall w,x,y,z\in N)\{[(w,x)\in P\wedge(w,y)\in D \wedge(x,z)\in D]\Rightarrow[(y,x)\in P]\}.$$

如果成分结构树中的标记 Q 是从语法范畴(如名词、动词、名词词组、动词词组等)和符号串元素(如自然语言中各个具体的单词)的有限集合中选出的,那么就可以用这种成分结构树来表示句子的层次结构及各成分从左到右的顺序.

层次分析法(hierarchical analysis method) 语言形式分析的一种方法. 一个复杂的语言形式,不能一下子就把它分析为若干个单词,而要按层次逐层地进行分析. 例如,语言成分 A 要按照如下图所示的步骤来分析:



不是把 A 一下子就分成 $A11,A121,A122,A211,A212,A221,A222$,而是先把 A 分成 $A1$ 和 $A2$ 两部分,然后把 $A1$ 分成 $A11$ 和 $A12$ 两部分,把 $A2$ 分成 $A21$ 和 $A22$ 两部分,$A12$ 又可再分为 $A121,A122$ 两部分,…,这样一直分析下去,一直分析到单词为

止. 通常把 $A1$ 和 $A2$ 称为 A 的直接成分,把 $A11$ 和 $A12$ 称为 $A1$ 的直接成分,把 $A121$ 和 $A122$ 称为 $A12$ 的直接成分. 这种依次找出语言格式直接成分的方法,称为层次分析法,亦称直接成分分析法.

乔姆斯基范式(Chomsky normal form) 上下文无关文法规则的一种表示方式. 美国语言学家乔姆斯基(Chomsky,N.)证明了:任何由上下文无关文法生成的语言,均可由重写规则为 $A\to BC$ 或 $A\to a$ 的文法生成,其中 A,B,C 是非终极符号,a 是终极符号. 具有这样的重写规则的上下文无关文法,它的推导树均可简化为二元形式,这样,就可以采用二分法来分析自然语言,并采用二叉树来表示自然语言的句子结构. 因此,上下文无关文法的重写规则 $A\to BC$ 或 $A\to a$ 便称为乔姆斯基范式.

自顶向下分析法(top-down analysis method) 一种语言形式分析算法. 是根据形式文法的重写规则,由根开始自顶向下地构造推导树,一直到得出语言中的合格句子为止的一种语言形式分析算法.

自底向上分析法(bottom-up analysis method) 一种语言形式分析算法. 是根据形式文法的重写规则,自叶开始逐级向上归约,直到构造出表示句子结构的整个推导树为止的一种语言形式分析算法.

依尔利分析法(Earley analysis method) 一种语言形式分析算法. 它交替地使用自顶向下及自底向上两种方法来分析语言. 首先自顶向下地预测某个语言成分的起点,找出起点后,再自底向上长成一个子树,如此交替地分析,直到长成表示语言句子结构的整个推导树为止. 这种分析法是由美国学者依尔利(Earley)提出的,故称为依尔利分析法.

自嵌入文法(self-embedding grammar) 上下文无关文法的一种类型. 如果在上下文无关文法中,存在着某一个非终极符号 A,具有性质

$$A\overset{*}{\underset{G}{\Rightarrow}}\varphi A\psi,$$

这里 φ 和 ψ 是非空符号串,G 表示上下文无关文法,$\Rightarrow$ 表示推导关系,那么这个文法就是自嵌入文法. 如果 G 是非自嵌入的上下文无关文法,那么由 G 生成的语言 $L(G)$ 就是有限状态语言. 如果 $L(G)$ 是上下文无关语言,那么当且仅当文法 G 是具有自嵌入性质的上下文无关文法时,$L(G)$ 才不是有限状态语言.

基本转换(basic transformation) 转换语法的基本概念之一. 上下文无关文法中的非终极符号 $Y_1,Y_2,\cdots,Y_r$,对于两个自然数 r, n $(n\leqslant r)$,存在自然数序列 $\beta_0,\beta_1,\cdots,\beta_k$ 与辅助符号串的词典 $V_p(Z_1,Z_2,\cdots,Z_{k+1})$ 中的序列,使得:

1. $\beta_0=0$, $k\geqslant 0$,对于 $1\leqslant j\leqslant k$,有 $1\leqslant\beta_j\leqslant r$.

2. 对于每一个非终极符号串 $Y_1,Y_2,\cdots,Y_r$,有

$$t(Y_1,Y_2,\cdots,Y_n;Y_n,\cdots,Y_r)$$
$$=Y_{\beta_0}{}^\frown Z_1{}^\frown Y_{\beta_1}{}^\frown Z_2{}^\frown Y_{\beta_2}{}^\frown\cdots{}^\frown Y_{\beta_k}{}^\frown Z_{k+1}.$$

这样一来,t 就把在上下文

$$Y_1{}^\frown Y_2\cdots{}^\frown Y_{n-1}-Y_n{}^\frown\cdots{}^\frown Y_r$$

中的 Y_n,转换成某个符号串

$$Y_{\beta_0}{}^\frown Z_1{}^\frown Y_{\beta_1}{}^\frown Z_2{}^\frown Y_{\beta_2}{}^\frown\cdots{}^\frown Y_{\beta_k}{}^\frown Z_{k+1},$$

这时 t 称为基本转换.

导出转换(derivational transformation) 转换语法的基本概念之一. 由基本转换 t 可得到导出转换 t^*. t^* 是基本转换 t 的导出转换,当且仅当对于一切的 $Y_1,Y_2,\cdots,Y_r$,有

$$t^*(Y_1,Y_2,\cdots,Y_r)=W_1,W_2,\cdots,W_r,$$

其中,对于每一个 $n\leqslant r$,都有

$$W_n=t(Y_1,Y_2,\cdots,Y_n;Y_n,\cdots,Y_r).$$

这样,导出转换 t^* 便把符号串 $Y_1,Y_2,\cdots,Y_r$ 转换成一个新的符号串 $W_1,W_2,\cdots,W_r$.

基本转换 t 和导出转换 t^*,构成了非终极符号串 $Y_1,Y_2,\cdots,Y_r$ 到另一个新的非终极符号串 $W_1,W_2,\cdots,W_r$ 的一个转换,这个转换记为 T.

转换模型(transformational model) 语言的一种形式模型. 转换模型由下面三个层级组成:

1. 直接成分层级. 在这个层级,利用上下文无关文法和上下文有关文法中的重写规则,生成语言核心句的非终极符号串 $Y_1,Y_2,\cdots,Y_r$,重写规则如下:

$$X_1\to Y_1$$
$$X_2\to Y_2$$
$$\cdots\cdots\cdots$$
$$X_r\to Y_r$$

2. 转换层级. 在这个层级,利用转换规则

$$T_1$$
$$T_2$$
$$\cdots$$
$$T_j$$

把核心句的非终极符号串 $Y_1,Y_2,\cdots,Y_r$,转换成另一个非终极符号串 $W_1,W_2,\cdots,W_r$.

3. 语素音位层级. 在这个层级,把转换所得到的非终极符号串 $W_1,W_2,\cdots,W_r$,按语言的语素音位规则转换为终级符号串 $w_1,w_2,\cdots,w_r$,

$$W_1\to w_1$$
$$W_2\to w_2$$
$$\cdots\cdots\cdots$$
$$W_r\to w_r$$

利用转换模型,可以把语言中的一些基本的句子作为核心句,由它们转换出该语言中数以万计的句子来,这样以简驭繁,便把语言的描写大大地简化了.

词域(realm of word)　用集合论方法描述语言的一个基本概念.某一个词的词域,也就是某一个词的词形变化的全部形式的集合.例如,英语动词learn(学习)的词域就是集合{learn, learned, learnt, learning}.一般把词 x 的词域记为 $\Gamma(x)$.

词的等价(equivalence of words)　用集合论方法描述词间关系的一个基本概念.如果两个词 x 与 y 满足下列条件,则称词 x 等价于词 y:

1.对于任何一个形如 A_1xA_2 的句子成立,句子 A_1yA_2 也成立.

2.对于任何一个形如 B_1xB_2 的句子成立,句子 B_1yB_2 也成立,其中 A_1, A_2, B_1, B_2 是任意的(可空)符号串.

词 x 与 y 等价常记为 $x \sim y$.

词族(family of words)　用集合论方法描述词的类聚的一个基本概念.按照词的等价关系把一个语言中的词划分为若干等价类,其中每个类称为一个词族.含有词 x 的词族通常记为 $S(x)$.例如,取英语中成立的句子:

1. This man is my teacher.
(这个人是我的老师.)

2. Who is that man over there?
(那边的那个男人是谁?)

在句子1中,词 man 以两个词串作为其环境,一个是 this,一个是 is my teacher, 在这个环境中,出现词 lady, woman, girl 等仍得成立句子;在句子2中,词 man 以词串 who is that 和 over there 为其环境,在这个环境中,出现词 lady, woman, girl 等仍得成立句子.因此,词 man, lady, woman, girl 等价,属于一个词族.

***B* 划分**(B-partition)　用集合论方法描述词的集合的基本概念之一.用彼此不相交子集 B_i 之并的形式来表示语言中词的集合 Ξ,即

$$\Xi = B_1 \cup B_2 \cdots \cup B_i \cdots \cup B_n = \bigcup_{i=1}^{n} B_i,$$

就把 Ξ 称之为集合的 B 划分.若 $x \in B_i$,有时可把 B_i 写为 $B(x)$,并有以下三种情形:

1.如果每一个子集只由一个词构成,就把这种划分称之为 E 划分.

2.如果每一个子集是词域,就把这种 B 划分称之为 Γ 划分.

3.如果每一个子集是词族,就把这种 B 划分称之为 S 划分.

***E* 划分**(E-partition)　见"B 划分".

Γ 划分(Γ-partition)　见"B 划分".

***S* 划分**(S-partition)　见"B 划分".

***B* 结构**(B-structure)　用集合论方法描述词的集合的基本概念之一.取任何一个句子 $A = x_1x_2\cdots x_i\cdots x_n$,则把子集

$$B(x_1)B(x_2)\cdots B(x_i)\cdots B(x_n)$$

的序列,即在给定的 B 划分中,词 x_i 所进入的子集的序列,称之为句子 A 的 B 结构,记为 $B(A)$.

取同一个句子 A = this man is my teacher 为例,在不同的 B 划分下,这个句子的 B 结构形式不同:

1.在 E 划分下,B 结构有形式

$E(A)$ = {this} {man} {is} {my} {teacher},

这种 B 结构可称为 E 结构.

2.在 S 划分下,B 结构有形式

$$S(A) = \begin{Bmatrix}\text{this}\\\text{that}\\\vdots\end{Bmatrix}\begin{Bmatrix}\text{man}\\\text{lady}\\\text{woman}\\\text{girl}\end{Bmatrix}\begin{Bmatrix}\text{is}\\\vdots\end{Bmatrix}\begin{Bmatrix}\text{my}\\\text{your}\\\text{his}\\\vdots\end{Bmatrix}\begin{Bmatrix}\text{teacher}\\\text{friend}\\\vdots\end{Bmatrix},$$

这种 B 结构可称为 S 结构.

3.在 Γ 划分下,B 结构有形式

$$\Gamma(A) = \begin{Bmatrix}\text{this}\\\text{those}\end{Bmatrix}\begin{Bmatrix}\text{man}\\\text{men}\end{Bmatrix}\begin{Bmatrix}\text{is}\\\text{are}\end{Bmatrix}\begin{Bmatrix}\text{my}\\\text{our}\end{Bmatrix}\begin{Bmatrix}\text{teacher}\\\text{teachers}\end{Bmatrix},$$

这种 B 结构可称为 Γ 结构.

如果至少有一个成立的句子具有某一 B 结构,那么这个 B 结构就是成立的.

***E* 结构**(E-structure)　见"B 结构".

***S* 结构**(S-structure)　见"B 结构".

Γ 结构(Γ-structure)　见"B 结构".

***B* 等价**(B-equivalence)　用集合论方法描述词的集合的基本概念之一.集合 B_i 与集合 B_j 如果满足下述条件,则称 B_i 与 B_j 是 B 等价的,记为 $B_i \sim B_j$:

1.对于任何一个形如 $B(A_1)B_iB(A_2)$ 的成立结构,结构 $B(A_1)B_jB(A_2)$ 也成立.

2.对于任何一个形如 $B(D_1)B_jB(D_2)$ 的成立结构,结构 $B(D_1)B_iB(D_2)$ 也成立.

词的等价是 B 等价的一种特殊情况,可称为 E 等价.B 等价是自反的、对称的、传递的,满足数学上等价关系的条件.

导出划分(derivational partition)　用集合论方法描述词的集合的基本概念之一.给定一个 B 划分 B,由相应的 B 等价可以产生词的集合 Ξ 的一个新的划分,称为 B 划分的导出划分,记为 B'.

由导出划分 B' 产生的新导出划分 B'' 与原导出划分 B 重合,即 $B''=B$.

词型(type of word)　词的集合的一种描述形式.Γ 划分的导出划分记为 T,即 $\Gamma' = T$.词 x 的词型记为 $T(x)$.如果两个词域彼此等价,它们就属于一个词型.例如,取这样的词域:

1.名词 table, tables 为一个词域,记为

Γ(table).

2. 名词 chair，chairs 为一个词域，记为

Γ(chair).

显然有

$$\Gamma(\text{table}) \underset{B}{\sim} \Gamma(\text{chair}),$$

因此，它们属于一个词型 T (table). 词 table 的一切形式，词 chair 的一切形式以及英语其他名词的一切形式都进入这个词型 T (table)之中，可见词型 T(table)很接近英语名词这个词类.

简单语言(simple language) 用集合论方法描述语言的基本概念之一. 如果在语言 Ξ 的词域与词族之间，满足如下两个条件，则称语言 Ξ 为简单语言：

1. 对于任何的 x，$\Gamma(x) \cap S(x) = x$ 成立，换言之，要求进入同一个词域中的两个词属于不同的词族.

2. 如果 $x' \in \Gamma(x)$，$y \in S(x)$，那么

$$S(x') \cap \Gamma(y) \neq \varnothing,$$

换言之，如果有词 x' 进入 $\Gamma(x)$ 中，又有词 y 进入 $S(x)$ 中，那么就应该存在某一个词 y' 进入 $S(x')$ 与 $\Gamma(y)$ 中.

***B* 构形**(*B*-configuration) 用集合论方法描述语言的基本概念之一. 取语言 Ξ 的任意 B 划分，如果下列条件成立，则把 $\widetilde{B}_{(1)}$ 称为一级 B 构形：

1. $\widetilde{B}_{(1)}$ 含有的元素不少于两个.

2. 存在着 B 划分的一个元素 B_{α_1}，使得 B 结构

$$\widetilde{B}_{(1)}B(A_2) \text{及} B(A_1)B_{\alpha_1}B(A_2)$$

在任何句子 A_1 与 A_2 中，同时成立或同时不成立.

元素 B_{α_1} 可以在保持结构成立的条件下替换构形 $\widetilde{B}_{(1)}$，称 B_{21} 是构形 $\widetilde{B}_{(1)}$ 的结果元. 结果元可以不是惟一的. 如果 B_{α_1} 是构形 $\widetilde{B}_{(1)}$ 的结果元，那么 B 划分中与 B_{α_1} 处于 B 等价的任何元素

$$B_i(B_i \underset{B}{\sim} B_{\alpha_1})$$

也可以是构形 $\widetilde{B}_{(1)}$ 的结果元. 用结果元 B_{α_1} 来替换一级 B 构形 $\widetilde{B}_{(1)}$，便得到一级 B 结构，记为 $B_{(1)}$.

在一般情况下，如果下列条件成立，则称 $B_{(n)}$ 为 n 级 B 构形：

1. $\widetilde{B}_{(n)}$ 含有的元素不少于两个.

2. 存在着一个元素 B_{α_n}，使得$(n-1)$级 B 结构

$$B(A_1)\widetilde{B}_{(n)}B(A_2) \text{及} B(A_1)B_{\alpha_n}B(A_2)$$

在任何句子 A_1 与 A_2 中同时成立或同时不成立，其中不包含 n 级 B 构形的 B 结构 $B(A_1)B_{\alpha_n}B(A_2)$ 称为 n 级 B 结构.

可见 B 构形的定义是递归的：通过$(n-1)$级 B 结构来定义 n 级 B 构形，通过$(n-2)$级 B 结构来定义$(n-1)$级 B 构形……，如此等等. 这样，每一个 B 构形用结果元替换之后就得到了同级的 B 结构. 根据 B 构形，可以把复杂的语言结构按层次一步一步地约化为简单结构.

***C* 等价**(*C*-equivalence) 词的等价关系的一种类型. 两个词 x 与 y，如果存在词的序列 $x_1, x_2, \cdots, x_n$，使得下列条件成立，则称词 x 与 y 为 C 等价的：

1. 对于任何的 i，x_i 与 x_{i+1} 或者属于同一个词域，或者属于同一个词族.

2. $x_1 = x$，$x_n = y$.

具有性质 1 的词的序列称为链(chain). 例如，下面的词构成一个链：

cat-cats-dogs,

其中，cat 与 cats 属于同一个词域，cats 与 dogs 属于同一个词族. C 等价也是一种等价关系，它可以得到集合 Ξ 上的一个新的划分. 由 C 等价得出的等价类称为节段.

语法性数学模型(mathematical model of grammatical gender) 用数学方法对语法性的形式描述. 选择两个名词 ξ 和 η 作为出发点，ξ 代表语义上的阳性，η 代表语义上的阴性. 例如，在法语中选择 garcon(男孩)为 ξ，选择 fille(女孩)为 η. 然后，以链的方式，把 Γ 划分与 S 划分交错地联系起来，从而来刻画自然语言的语法性. 这样的数学模型称为语法性数学模型.

一个名词属于语法阳性，则对于任何的 $x' \in \Gamma(x)$ 和任何的 $\xi' \in \Gamma(\xi)$，都可以用一条长度不大于 3 的链连结起来.

一个名词属于语法阴性，则对于任何的 $y' \in \Gamma(y)$ 和任何的 $\eta' \in \Gamma(\eta)$，都可以用一条长度不大于 3 的链连结起来.

既不是阳性又不是阴性的名词，就是中性名词.

既是阳性又是阴性的名词，就是双性名词.

例如，在法语中有

$$\xi = \text{garcon}, \quad \eta = \text{fille},$$

可得到如下的链：

$$\text{femmes} \xrightarrow{\Gamma} \text{femme} \xrightarrow{S} \text{fille},$$

$$\text{chambres} \xrightarrow{\Gamma} \text{chambre} \xrightarrow{S} \text{fille},$$

$$\text{maisons} \xrightarrow{\Gamma} \text{maison} \xrightarrow{S} \text{fille},$$

以及

$$\text{hommes} \xrightarrow{\Gamma} \text{homme} \xrightarrow{S} \text{garcon},$$

$$\text{crayons} \xrightarrow{\Gamma} \text{crayon} \xrightarrow{S} \text{garcon},$$

$$\text{murs} \xrightarrow{\Gamma} \text{mur} \xrightarrow{S} \text{garcon}.$$

按定义可推出 femme(女人)，chambre(房间)，maison(屋子)为阴性；homme(男人)，crayon(铅笔)，mur(墙)为阳性.

句法类型(syntactic types)　范畴语法的基本概念之一.任何词都可以根据它在句子中的功能归入一定的句法类型.如果用 n 表示名词的句法类型,用 S 表示句子,则其他的一些句法类型都可以用 n 和 S 以不同的方式结合起来表示.规则是:

1.如果有某个词 B,其后面的词 C 的句法类型是 γ,而它们所构成的词的序列 BC 的功能与 β 相同,则这个词 B 的句法类型记为 β/γ.

2.如果有某个词 B,其前面的词 A 的句法类型为 α,而它们所构成的词的序列 AB 的功能与 β 相同,则这个词 B 的句法类型记为 $\alpha\backslash\beta$.

3.如果有某个词 B,其前面的词 A 的句法类型为 α,其后面的词的句法类型为 γ,而它们所构成的词的序列 ABC 的功能与 β 相同,则这个词 β 的句法类型为 $\alpha\backslash\beta/\gamma$.

根据这种记法,可以写出自然语言中词的句法类型.例如,在英语中,John 的句法类型为 n; poor John 中的 poor,它后面出现名词 John,而它所构成的 poor John,功能与名词相同,故其句法类型为 n/n; John works 中的 works,它前面出现名词 John,而它所构成的 John works,功能与句子相同,故其句法类型为 $n\backslash S$. John likes Jane 中的 likes,它前面为名词 John,后面为名词 Jane,而它所构成的 John likes Jane 功能与句子相同,故其句法类型为$n\backslash S/n$.

句法类型演算(syntactic types calculus)　范畴语法的一种演算方法.对于语言中的词列出了一个完整的句法类型清单之后,便可根据如下规则进行句法类型演算:如果有形如 $\alpha,\alpha\backslash\beta/\gamma,\gamma$ 的符号序列,那么就用 β 来替换它.这个规则同时也包括下面两个规则:

1.用 β 替换形如 $\alpha,\alpha\backslash\beta$ 的符号序列.

2.用 β 替换形如$\beta/\gamma,\gamma$ 的符号序列.

如果把语言中的词标上句法类型,通过有穷个演算步骤,可把词的序列化为 S,则这个词的序列便是该语言中的合格句子.例如,

John	likes	Jane
n	$n\backslash S/n$	n
	$n\backslash S$	
	S	

蒙塔鸠语法(Montague grammar)　建立在数理逻辑基础上的形式语法.它是采用内涵逻辑的方法来描述句子语义内容的一种理论.1970 年前后,美国数理逻辑学家蒙塔鸠(Montague, R.)等人把内涵逻辑应用于自然语言的研究,并把生成语法与内涵逻辑这两个领域的研究集中提炼为蒙塔鸠语法.蒙塔鸠语法把用树形图表示的句子的深层结构转换为内涵逻辑表达式,从而揭示句子的某些语义内容.这种语法理论对于自然语言信息处理、逻辑学,以致于认知科学是很有价值的.

复杂特征集(complex features)　数理语言学的基本概念之一.用来描述自然语言中的语言单位的语法语义类别、功能和属性的复杂标记的集合.通常记为

$$\begin{bmatrix} f_1 = v_1 \\ f_2 = v_2 \\ \vdots \\ f_n = v_n \end{bmatrix} \quad (n \geqslant 1),$$

其中 f_i 为特征名,v_i 为特征值,f_i 是一个原子,v_i 是一个原子或者一个复杂特征集.例如,英语中作 see 的过去时用的语言单位 saw,可用复杂特征集描述如下:

$$\begin{bmatrix} \text{category}=\text{verb} \\ \text{tense}=\text{past} \\ \text{transitivity}=\text{mental}-\text{process} \\ \text{root}=\text{see} \\ \text{lex}=\text{saw} \end{bmatrix}.$$

这种复杂特征集是采用多标记来对自然语言的单位进行描述.中国学者在 20 世纪 80 年代初期,在机器翻译研究中就是采用多标记的树形图来描述源语言和目标语言的,这种多标记的集合实际上就是复杂特征集.

描述元(descriptor)　数理语言学的基本概念之一.复杂特征集中的一个成分集、一个模式或一组带属性的值,其中最主要的是"属性/值"偶对,构成一个描述元,记为 $f_i=v_i$.

合一运算(unification)　自然语言处理中复杂特征集的运算.如果运算 $\Cup$ 具有如下性质,则 $\Cup$ 是合一运算:

1.若 a 和 b 均为原子,则 $a\Cup b=a$,当且仅当 $a=b$;否则 $a\Cup b$ 为空.

2.若 α 和 β 均为复杂特征集,则:

1) 若 $\alpha(f)=v$,但 $\beta(f)$的值未经定义,则 $f=v$ 属于 $\alpha\Cup\beta$.

2) 若 $\beta(f)=v$,但 $\alpha(f)$的值未经定义,则 $f=v$ 属于 $\alpha\Cup\beta$.

3) 若 $\alpha(f)=v_1,\beta(f)=v_2$,且 v_1 与 v_2 不相抵触,则 $f=(v_1\Cup v_2)$属于 $\alpha\Cup\beta$,否则 $\alpha\Cup\beta$ 为空.

集合论中的求并运算是合一运算的一种特殊情况.当合一的对象所含的元素为不可分解的原子时,合一的结果等于并集.当合一的对象是有结构的复杂特征集时,就要检验特征的相容性.只有当特征相容时,相应的复杂特征才能合一,如果参与合一的特征相冲突,就立即宣布合一失败.合一运算在合并自然语言处理中各方面来的特征信息的同时,还提供

了检验限制条件的机制.

n 元语法(n-gram grammar) 建立在马尔可夫模型上的一种概率语法. 它通过对自然语言的符号串中 n 个符号同时出现概率的统计数据来推断句子的结构关系. 当 $n=2$ 时,称为二元语法,当 $n=3$ 时,称为三元语法.

统计语言学

统计语言学(statistical linguistics) 数理语言学的一个分支学科. 它是应用统计方法来研究语言现象的学科. 其内容主要包括以下几个方面:

1. 统计语言单位的出现频率.

2. 统计作家的用词频率、词长分布和句长分布,以了解作家运用语言的风格.

3. 计算语言存在的绝对年代以及亲属语言从共同原始语分化出来的年代,对亲属语言的语法、语音体系进行统计和比较.

4. 研究语言的熵和羡余度.

5. 探讨自然语言的一般统计规律.

6. 运用马尔可夫过程的理论来研究语言.

7. 研究文章中词与词之间、两个语法范畴之间、两个语义类别之间或两个句法类型之间的间距,以揭示文章的句法和语义特征.

8. 研究语言的词汇与文章长度的关系,以揭示文章中词汇的丰富程度和差异程度.

9. 研究词典的数学结构.

统计语言学在数理语言学研究中有比较长的历史,它的研究成果对于通信技术、语言教学和自然语言的信息处理都很有价值.

频率词典(frequency dictionary) 词典的一种类型. 它是按词出现频率的高低而编排的词典. 世界上第一部频率词典是 1898 年德国语言学家凯定(Kaeding, F. W.)编的《德语频率词典》,这部词典统计了 110 万个词的语言素材.

频率词典按其内容可分为普通频率词典和专业性频率词典两大类. 普通频率词典又可分为单语言频率词典和多语言频率词典. 专业性频率词典按不同专业中单词出现频率的大小来编排.

齐普夫定律(Zipf law) 统计语言学中的一个重要定律. 它是美国语言学家齐普夫(Zipf, G. K.)提出的一个经验公式. 在按频率递减顺序排列的频率词典中,词的序号 r 和具有该序号的词的频率 f_r 之间,存在着关系

$$f_r=cr^{-1},$$

其中 c 是一个常数,其值的区间为 $0<c<0.1$.

朱斯定律(Joos law) 统计语言学中的一个重要定理. 指美国语言学家朱斯(Joos, M.)对齐普夫定律的一种修正. 他用 p_r 表示频率词典中序号为 r 的词的频率,则有

$$p_r=cr^{-b},$$

其中 $b>0, c>0$. 朱斯公式中有 b 和 c 两个参数,因此又称为双参数序号分布定律. 齐普夫公式是朱斯公式中当 $b=1$ 时的一种特殊情况,它是单参数序号分布定律.

曼德勃罗特定律(Mandelbrot law) 英籍法国数学家曼德勃罗特(Mandelbrot, B. B.)对齐普夫定律的一种修正. 他利用概率论和信息论方法研究词的序号分布规律,他把空白看成是以空白为结尾的字母的随机序列,又把句子看成是用词来编码的词的序列,把文章看成是由句子的变化过程而形成的句子的序列,通过严格的数学推导,提出了词的三参数序号分布定律,故称其为曼德勃罗特定律,其形式为

$$p_r=c(r+a)^{-b},$$

其中 $0\leqslant a<1$, $b>0, c>0$. a, b, c 三个参数的含义如下:

1. 参数 c 与出现概率最高的词的概率的大小有关.

2. 参数 b 与高概率词的数量的多少有关,对于 $r<50$ 的高概率词,b 是 r 的非减函数,随着 r 的增大,参数 b 并不减小.

3. 参数 c 与词的数量 n 有关,由于 a 的选择自由度较大,因而公式的灵活性很大,能在各种情况下适合测定的数据.

在曼德勃罗特公式中,当 $a=0$ 时,公式形式为

$$p_r=cr^{-b},$$

这就是朱斯公式;当 $a=0, b=1$ 时,公式形式为

$$p_r=cr^{-1},$$

这就是齐普夫公式. 可见,朱斯公式和齐普夫公式只不过是曼德勃罗特公式的特殊形式.

FEL 公式(FEL formula) 用于描述术语数据库的一个公式. 在一个术语数据库或者术语词典中,术语系统的经济指数 E 与术语的平均长度 L 的乘积等于单词的术语构成频率 F,可写为

$$F=EL,$$

这个公式称为 FEL 公式.

术语系统的经济指数 E 表示单词构成术语数量的多少,单词构成的术语数量越多,术语系统的经济指数越大;术语的平均长度 L 表示术语系统中平均每个术语所包含的单词数目的多少,术语中包含的单词越多,术语的平均长度越大;单词的术语构成频率 F 表示术语系统中高频词含量的多少,术语系统中的高频词越多,则该系统中单词的术语构成频率就越高. FEL 公式反映了术语系统中这三个量之

间的函数关系.

术语形成的经济律(economical law for term formation)　FEL公式所反映的术语形成规律.根据FEL公式,在不改变术语平均长度的条件下,为了提高术语系统的经济指数,就要提高单词的术语构成频率,也就是用有限数量的单词来构成大量的术语.这样,在术语的形成过程中,有限数量的单词将会产生大量的词组型术语,使得词组型术语的数量大大地超过了单词型术语的数量,而成为术语系统中的大多数.这说明,在术语形成过程中存在着用有限的单词来构成无限的术语的规律.术语系统中单词的数目是有限的,而由单词构成的词组型术语的数目却是无限的.术语系统的形成过程就是有限手段的无限运用的过程,它反映了术语形成的经济规律,这个规律称为术语形成的经济律.语言是现实的编码体系,术语形成的经济律可以推广到语言编码的其他领域.

汉语中在用单字组成多字词的时候,有限数目的单字组成了为数可观的多字词,多字词以增加自身字的长度(即多字词中包含汉字数目的多少)为代价来保持汉语中原有的单字的个数或者尽量不增加原有单字的个数,体现了组字成词这一编码过程的经济律.

对汉字结构及其构成成分的统计与分析表明:在《辞海》所收的16295个汉字和《国家标准信息交换用汉字编码字符集·基本集》收入而《辞海》未收的43个字中,简化字和被简化的繁体字以及未简化的汉字共有16339个,它们是由675个不能再分解的末级部件构成,简化字和未简化的汉字共11837个,它们是由648个不能再分解的末级部件构成.由少量的部件构成大量的汉字,体现了部件构成汉字这一编码过程的经济律.

如果把术语形成的经济律进一步推广到组字成词和由部件构成汉字等语言编码的领域,这个经济律就成为具有一般性的"语言编码的经济律".

尤尔图(Yule graph)　一种估计不同作家风格的统计特征的方法.是英国数学家尤尔(Yule,G. U.)提出的.研究任何一个作家包含n个词的文章,设有n个词在文章中出现一次,其出现概率为n_1/N,有n个词在文章中出现两次,其出现概率为$n_2\times 2/N$.一般地,设有n个词在文章中出现i次,其出现概率为

$$p(i)=n_i\times\frac{i}{N}.$$

在直角坐标系中,做出i与$p(i)$的关系曲线,这个曲线就称为尤尔图.比较不同作家的尤尔图,就可以对他们各自的文章中词的分布特征做出某些结论.利用尤尔图还可以确定匿名文章的作者是谁,这对于古籍考证和侦察工作都是有价值的.

富克斯公式(Fuchs formula)　描述词的音节数的一个公式.富克斯(Fuchs,H. M. L.)用概率方法研究不同语言中词的音节数目的分布规律,得到了如下的公式:

$$p_i=\frac{e^{-(\bar{i}-1)}\cdot(\bar{i}-1)^{i-1}}{(i-1)!},$$

式中i表示词中的音节数($i=1,2,3,\cdots$),$\bar{i}$表示某一语言中词的音节数的平均数,e为自然对数的底数,p_i表示有i个音节的词在文章中的百分比.这个公式称为富克斯公式.利用这个公式,可以根据词的音节数及词的音节数的平均数,直接计算出具有某一音节数的词的百分比.

语言成分熵(entropy of language elements)　语言成分信息量的数学量度.从信息论的角度来看,使用自然语言来交际的过程,也就是从语言的发送者通过通信媒介传输到语言的接收者的过程.在语言的接收者接收到语言消息之前,他对该消息的不定度是很大的,这种不定度的大小称为语言成分的熵.在接收到语言成分之前,熵因语言成分符号的数目和出现概率的不同而不同,在接收到语言符号之后,不定度被消除,熵等于零.

在考虑语言成分出现概率差异的情况下,语言成分的熵可以用下面的公式来计算:

$$H_1=-\sum_{i=1}^{n}p_i\log_2 p_i,$$

其中,H_1是熵,p_i是语言成分的出现概率,n是该语言中语言成分数目.因此,英语按字母计算的熵约为4.03bit,法语按字母计算的熵约为3.98bit,德语按字母计算的熵约为4.037bit,俄语按字母计算的熵约为4.35bit.

汉字数目约8万个,因而汉字熵的计算极为困难.中国学者提出了"汉字容量极限定律",从理论上证明了:在考虑汉字出现概率差异的情况下,当汉字容量小于12366时,汉字的熵随着汉字容量的增加而增加,当汉字的容量大于12366时,汉字的熵就不再增加而稳定于9.65bit.当考虑汉字彼此之间的影响时,还可以计算语言成分的极限熵.中国学者根据英汉双语语料库的文本容量对比,计算出汉字的极限熵介于3.0212bit与5.0713bit之间,其平均值为4.0462bit,这个极限熵的值就是人们读到一个汉字时获得的实际信息量.

语言羡余度(redundancy of language)　语言成分羡余量的数学描述.由于语言的结构而产生的语言中多余成分的百分比.可用下述公式计算:

$$R=1-\frac{H_\infty}{H_0},$$

其中 R 表示羡余度，H_∞ 是极限熵，H_0 是当语言中各成分都是等概率不相关时的熵.

语言中的羡余度有四种类型：

1. 在书面语中，有不少字母是由于语言结构规定好的，根据前面出现的字母，往往可以预测后面的字母是什么.

2. 在口语中，当人们某些音发得不清楚时，听话人根据口语的语境仍然可以听懂.

3. 在文字中，并非构成字母或汉字的一切笔画对于辨识这个字母都是必不可少的；在一行印刷的字母或汉字中，遮住其下面的一半，仍然能毫不困难地把这一行读懂.

4. 在某个音素中，并非语音的一切要素（音强、音高等特征）对于辨识这个音素都是必不可少的. 语音似乎穿着许多件松软的"外衣"，在无损于人们辨识它的条件下，剥去几件"外衣"也是可以的.

这四种类型的羡余度都是有益的，它保证语言在不理想的条件下（如书面文句有遗漏，谈话时有嘈杂声，书写的字母不清楚，发音不清楚），仍然能够发挥其交际功能. 但是，当在通信中沿信道传输语言信息时，羡余度会使信道负荷过重，要通过适当的编码方式来降低羡余度.

语言年代学（glottochronology） 数理语言学的分支学科. 是美国语言学家斯瓦德士（Swadesh, M.）提出的一种测定古代语言年代的学说. 斯瓦德士指出，每一种语言都有一些基本词汇，如人称代词、身体各个部分的名称等，这些基本词汇的变化速度，在很长的时间内大体上是一样的. 他选择 200 个词作为适用于各种语言的基本词汇，经过统计计算，得出它们在 1000 年中保存下来的词汇大约为 86%. 如果某种古代语言及其发展而成的现代语言的基本词汇有 60%是相同或相近的，那么可根据公式

$$t=\frac{\ln l}{\ln l_0}$$

来计算这种古代语言存在的绝对年代，其中 l_0 等于 0.86，l 是在该现代语言中保留下来的基本词汇的百分比，t 是该古代语言存在的绝对年代.

根据条件 $l=0.60$，故得

$$t=\frac{\ln l}{\ln l_0}=\frac{\ln 0.60}{\ln 0.86}\approx 3(\text{千年}).$$

也就是说，这种语言从古代算起已经存在 3000 年了.

如果比较的不是古代语言及其发展而成的现代语言，而是两种由共同原始语分化而来的现代语言，要使这两种现代语言的基本词汇中共同的词的比例为 L，那么这两种现代语言从原始语分化的绝对年代可按公式

$$t=\frac{\ln L}{2\ln l_0}$$

来计算. 例如，比较英语和德语的基本词汇得出：$L=0.82$，由此可算出

$$t=\frac{\ln L}{2\ln l_0}=\frac{\ln 0.82}{2\ln 0.86}\approx 1.3(\text{千年}).$$

这意味着，英语和德语是在 1300 年前，即公元 6 世纪时分化的.

语言演变的因素是比较复杂的，民族迁徙、民族接触以及其他社会历史因素，经常会加快或减慢语言词汇的变化速度. 语言年代学没有考虑到这些复杂因素，有一定的局限性.

计算语言学

计算语言学（computational linguistics） 计算机科学与语言学的交叉学科. 它是利用计算机对自然语言进行研究的一个新兴学科，始于 20 世纪 50 年代的机器翻译研究. 计算语言学包括计算语音学、计算词汇学、计算语法学、计算语义学等分支学科. 计算语音学研究如何用计算机对语音信息进行处理，实现语音的自动合成与分解；计算词汇学研究如何利用计算机处理自然语言的词汇，建立语言词汇库、术语数据库等机器可读词典；计算语法学研究如何用计算机来分析自然语言的语法，这种研究称为自动语法分析；计算语义学研究如何用计算机来分析自然语言的语义. 语义的形式化比较困难，目前正在探讨把语法的形式化与语义的形式化结合起来的有效方法.

计算风格学（computational stylistics） 统计语言学的分支学科. 它是采用计算机技术和统计方法，编制特定的程序和数学模型来研究自然语言的文章风格的学科. 主要研究内容有：

1. 根据语言单位之间的数量关系，确定文章风格的差异.

2. 把语言成分看做随机变量，研究语言变量之间的相关程度，推断风格的变异情形.

3. 把语言集团的语言变异与语言集团的地理位置结合起来研究，测定不同地理区域的人群之间的语言关系，确定相关性最高的区域.

4. 通过语言单位出现频率的统计，揭示语言成分在较大序列中的分布特征，了解不同文间的风格，判断文章的作者或写作年代.

汉字编码技术（Chinese character coding technique） 中文信息处理的一种技术. 它是研究汉字编码字符集的设计和汉字编码键盘输入的设计的信息技术. 它涉及汉字的收字、分级、排序、定位、认同等规则，简体字、繁体字和异体字关系的研究与确

定,建立字符集及其编码之间的映射关系,构造完备的计算机汉字代码系统以及汉字编码键盘输入系统等.

机器翻译系统(machine translation system) 计算语言学的一个研究领域.它是实现机器翻译的计算机系统,它通常包括机器词典、句法语义分析规则、原语译转换规则、译语生成规则以及翻译支援系统等.

汉语自动切词(automatic segmentation word of Chinese) 中文文本处理的一种技术.把输入计算机的汉语语句自动切分为单词系列的过程,使单词与单词之间出现空白,以供进一步的自动处理.在特定情况下切词的结果也可以包括一些词组或语素.汉语自动切词的方法有最大匹配法、逆向最大匹配法、切分标志法、逐词匹配法、邻接表法等.

扩充转移网络语法(augmented transition network grammar) 简称ATNG.一种用于自然语言处理的形式语法.美国学者伍兹(Woods, W.)提出的.转移网络是一种用图来表达语法规则的形式化体系.如果对这种转移语法加以扩充,增加可用于存储中间结果的寄存器,并在每条弧上附加任意的测试和动作,就得到扩充转移网络,简称ATN.

定子句语法(definite clause grammar) 简称DCG.一种逻辑语法.美国学者佩瑞拉(Pereira, F.)等提出的用一阶谓词中的定子句来描述自然语言.由于定子句表达的语法规则本身就是逻辑程序设计语言Prolog的可执行程序,因而可以被Prolog系统直接解释.

词汇功能语法(lexical functional grammar) 简称LFG.一种形式语法.是美国计算语言学家布列斯南(Bresnan, J.)和卡普兰(Kaplan, R.)提出的一种用于自然语言处理的形式语法.词汇功能语法对于一种语言的描写由短语结构规则和说明短语句法功能的功能方程两部分组成.它对于一个句子的描写由成分结构和功能结构组成.成分结构相当于一个语法树,树上的结点用短语标记来标注.功能结构相当于一个复杂特征集.功能结构由功能方程生成,用于描写句法树上每个结点的语法功能和谓词变元关系.词汇功能语法重视词汇信息的描述,整个语法由词汇来驱动.

功能合一语法(functional unification grammar) 简称FUG.一种形式语法.是美国计算语言学家凯依(Kay, M.)提出的用于自然语言处理的形式语法.在这种语法中,词条定义、句法规则、语义规则以及句子的结构功能关系全部都由复杂特征集来表示,并采用合一运算的方法来进行运算.这种语法既可用于自然语言的自动分析,又可用于自然语言的自动生成,是一种双向性的语法.

广义短语结构语法(generalized phrase structure grammar) 简称GPSG.一种形式语法.是英国语言学家盖兹达(Gazdar, G.)等提出的一种用于自然语言处理的形式语法.在这种语法中,规则系统要经过一系列的合格性条件检验,才能跟句子的表层结构联系起来.这些合格性条件包括特征限制、隐含特征规定、中心语特征、基础特征原则、控制一致原则、线性前置陈述.这些合格性条件检验有效地改进了短语结构语法,提高了语法理论对于语言事实的解释能力.

语料库语言学(corpus linguistics) 研究语料库的一门新兴学科.它是通过对大量的真实语言材料的分析和统计来建立自然语言处理规则的一种语言学说,其目标是处理大规模的自然语言真实文本,并从语料中提取知识.

格语法(case grammar) 一种语法理论.是美国语言学家菲尔墨(Fillmore, C.J.)提出的.这种语法认为,在没有格形态的语言中,也存在施事、受事、使役、给予、工具等格关系,这样便为各种自然语言的形式描述提供了普遍而有效的理论模型.

国际计算语言学委员会(International Committee on Computational Linguistics) 简称ICCL.计算语言学工作者的国际性学术组织.成立于1961年,首任主席是法国数学家沃古瓦(Vauquois, B.),其宗旨是促进计算语言学领域内的国际学术交流.自1965年起,每两年召开一届国际学术会议,称为COLING.自1982年开始,中国学者参加了这个国际组织的学术活动.

撰　稿　冯志伟
审　阅　马希文　许孔时

力　　　学

力学(mechanics)　研究客观物质机械运动的学科. 机械运动是指物质在空间中的位置随时间的变化,静止和平衡是它的特殊情况,它是自然界存在的最普遍、最基本的变化形态. 物理学也研究客观世界的机械运动,而且在物理学发展的早期,力学是物理学的一个分支,以后随着历史的进程,力学和物理学的研究范围有了相对的分工. 力学侧重于研究宏观物体的机械运动,而物理学侧重于研究微观世界的机械运动,如原子、电子等的机械运动. 当然物理还研究除机械运动之外的一些变化,如电、热等. 由于机械运动的普遍性,自然界的其他运动形态,如化学、生物学、天文学中研究的对象都或多或少地涉及机械运动,因此力学和其他学科形成许多交叉学科,如天体力学、地质力学、生物力学等. 工程技术中有大量的力学问题,如近代的土木工程、机械工程、航空工程、航天工程等都是在力学理论的指导下发展起来的,所以力学又是现代工程技术的理论基础之一.

17 世纪,英国数学家、物理学家牛顿(Newton, I.)系统地总结了前人研究的成果,建立了自由质点运动的三大定律,并成功地解释了行星的运行,使力学成为自然科学最早精确化的学科. 后人将基于牛顿三定律建立起来的力学理论体系称为牛顿力学. 随后,法国数学家和力学家拉格朗日(Lagrange, J.-L.)、英国数学家和物理学家哈密顿(Hamilton, W. R.)利用分析的方法处理带约束条件的力学系统,发展了分析力学;瑞士数学家欧拉(Euler, L.)、法国数学家和工程师纳维(Navier, C.(-L.-M.-H.))、法国数学家和力学家柯西(Cauchy, A.-L.)、法国数学家和力学家泊松(Poisson, S.-D.)、英国数学家和天文学家斯托克斯(Stokes, G. G.)等发展了弹性力学和流体力学,从而奠定了连续介质力学的基础. 19 世纪末到 20 世纪,力学得到进一步发展,形成了众多的分支学科和庞大的理论体系. 力学是自然科学中最早、也最充分地精确化的学科. 力学的发展和数学的发展有着密不可分的联系,历史上许多伟大的力学家同时也是伟大的数学家,许多力学的划时代的成就伴随着数学上划时代的突破. 例如,牛顿力学的产生和微积分的诞生是同时的,分析力学原则的确立和 n 维空间的引进是同时的,变分法和力学中的变分原理则是同一命题的两种不同的表述形式等.

力学研究方法可以归结为理论方法和实验方法,近年来,由于电子计算机的发展又产生了计算方法. 相应于后两种方法,产生了实验力学和计算力学两个分支学科. 力学的分支学科可以从不同的侧面来划分. 以研究方法来划分,可分为理性力学、实验力学和计算力学,实验空气动力学、实验应力分析等又可看做实验力学的分支学科. 依研究对象来划分,可以分为一般力学、固体力学、流体力学、岩土力学等,固体力学又可分为弹性力学、塑性力学、断裂力学等,而流体力学又可分为空气动力学、多相流体力学、湍流、非牛顿流、渗流力学等. 依工程技术应用为专门对象,可分为水力学、爆炸力学、弹道力学、陀螺力学、复合材料力学、工业和环境空气动力学、结构力学等. 按教学需要又分为理论力学、工程力学、材料力学以及有关的专门力学等.

力 学 总 论

理性力学(rational mechanics)　力学的分支学科. 指以建立力学对象的公理模型、运用演绎推理方法研究力学规律的分支. 理性力学一词最早由 18, 19 世纪, 法国数学家、力学家达朗贝尔(d'Alembert, J. le R.)等所引用,主要研究质点和刚体力学的内容,并建立了此部分问题的公理化体系. 20 世纪 50 年代以后,不少学者用公理化方法研究连续介质力学,并取得了相当大的进展. 这种方法已深入到热力学和其他连续体物理学科.

一般力学(general mechanics)　力学的分支学科. 主要研究对象为质点、刚体及由其组成的离散系统的力学行为,也研究包括离散系统、连续体、电磁系统和控制系统在内的复杂系统的动力学行为. 一般力学研究的问题常可用常微分方程组描述,所以,在数学上它与常微分方程理论和动力系统理论往往有共同的研究内容,如研究动力系统的可积性条件、运动的稳定性、运动的定性性质和极限性质等.

一般力学研究力学现象普遍性质的部分有:研究等价性原理和各种守恒规律的不变量理论,研究用广义坐标和相空间运动方程的表述形式及其解的性质的分析力学. 一般力学与工程技术有紧密联系的部分有:陀螺力学、振动理论、机器人动力学、多刚体系统动力学、变质量力学、弹道学等. 一般力学与其他学科交叉产生的方向有:天体力学、体育运动力学等. 连续介质力学在早期同一般力学有严格的界线,即前者研究连续介质,而后者研究离散体. 但由

于近年来的发展，这方面的分工变得不明显了．这是由于：一方面采用离散化方法，连续介质可以近似地用一个离散系统去描述；另一方面，有限自由度的某些一般性规律正在向无限自由度推广，开始了无限自由度和无限维动力系统的研究．

连续介质力学(mechanics of continuous media)　以连续介质理论模型为研究对象的力学分支．连续介质是客观物质中宏观物理量(如密度、速度、位移、压强、温度等)是空间和时间的连续函数的模型，所以在研究时可以对这些物理量使用数学分析的所有方法，如使用微分、积分、变分等数学工具．严格地讲，物体无限细分时，最终会达到不连续的微粒，如分子、原子．但是，连续介质力学讨论的是物体的宏观性质，所以可以把分子、原子的一个适当大的微团的平均物理量当做一点的物理量，这时物理量可以看成连续的，而且近似程度相当高．弹性体、流体、塑性体都是连续介质，所以通常的弹性力学、流体力学、塑性力学都是连续介质力学的分支．对于非常稀薄的气体，或由颗粒组成的固体，它们的物理量不再是坐标的连续函数，这时就不能再使用连续介质模型．

流体力学(fluid mechanics)　连续介质力学的一个分支．是研究流体的宏观平衡和流体运动规律以及与其他运动形态之间相互作用的学科．流体力学不考虑流体的微观结构，而把流体看做是在空间连续分布的连续介质，因而也是连续介质力学的一个分支．流体力学的基本内容可分为流体静力学、流体运动学和流体动力学．研究流体力学的方法有理论分析方法、数值计算方法、实验室模拟方法和现场观测方法．以流体力学为基础，已发展并建立起许多新的学科或分支，例如气体力学、水动力学、渗流力学、物理化学流体力学、磁流体力学、计算流体力学、生物流体力学、地球流体力学、爆炸力学等．

固体力学(solid mechanics)　连续介质力学的一个分支．是研究在外部因素(外力、温度、电磁场等)作用下，固体的变形及内部质点的位移、应力、应变和破坏等规律的学科．所谓固体是指客观连续物体中在受剪应力作用时可以维持静止状态的物体．固体物体，由于它们的本构关系不同，又可分为弹性体、塑性体、弹塑性体、粘性体、弹粘塑性体等，相应地，研究这些不同固体力学性质的学科称为弹性力学、塑性力学等．

对于不同几何形状和受力特点的固体，其近似处理方法也不同．在工程结构中常见的有杆与杆系、梁、板与壳体、薄壁杆件及块状结构等．固体力学不仅研究这些不同对象的平衡，即静力学问题，还研究它们的动力学问题，如振动问题、稳定性问题等．固体在一定条件下会发生破坏、断裂等现象，固体力学还研究这些现象的产生与发展，研究这些内容的分支称为材料强度理论、断裂力学、疲劳和累积损伤等．用多种材料造成的复合材料是近年来比较活跃的一个技术领域，复合材料力学是研究这种材料力学行为的分支．许多固体材料在外力作用下产生随时间缓慢的变形，称为流变性质，严格说来，它们已属于流体，所以研究这种流变固体时，固体力学和流体力学没有严格的界限．

材料力学(mechanics of materials or strength of materials)　固体力学的分支．一般是指研究固体材料的力学行为的学科．它研究固体材料的弹性、塑性、硬度、韧性、脆性、在常温和高温下的蠕变，研究在外部条件下的破坏准则、长期工作时的材料疲劳和损伤、裂纹对强度的影响及其扩展规律等，也研究材料力学性能的检验和有关标准，以及加工过程对材料力学性能的影响．

材料力学又是一门课程，指在工程教育中将有关固体力学最基础的内容组织起来的教学系统．通常包括杆与梁的理论和利用某些简化假定近似估算零件某些部位应力的方法、强度理论和某些工程专业的特别专题，如薄壁杆件、弹性基础上的梁、压力容器的强度等．

弹性力学(elasticity)　亦称弹性理论．固体力学的一个重要分支．是研究弹性体在外力、温度等作用下产生应力、应变和位移的学科，有时还要讨论弹性体在这些外因作用下平衡或运动的稳定性．许多工程材料，如金属、混凝土等所构成的物体在一定条件下都可以近似看做弹性体，所以弹性力学在工程中有很多应用．弹性力学问题可以表述为一组偏微分方程的初边值问题，求解它要用到较多的数学知识，所以有时又称为弹性数学理论．

弹性力学是连续介质力学中、也是固体力学中发展比较早的分支，它的许多基本概念同时也是连续介质力学和固体力学发展的基础．17世纪，英国自然科学家胡克(Hooke，R.)提出了线性弹性材料的早期模型；随后，法国数学家拉梅(Lamé，G.)和英国数学家格林(Green，G.)把它推广到一般情形；19世纪，法国数学家、力学家柯西(Cauchy，A.-L.)引进了应力和应变张量；同时，法国数学家、工程师纳维(Navier，C.(-L.-M.-H.))和拉梅等确定了弹性力学的基本方程．弹性力学中许多特殊问题的研究与它的一般理论研究并行发展着．早在17世纪，意大利科学家伽利略(Galilei，G.)就开始了梁的研究；18—19世纪，瑞士数学家约翰第三·伯努利(Bernoulli，Johann Ⅲ)研究并建立了梁的弯曲理论；德国学者基尔霍夫(Kirchhoff，G.R.)和纳维等在18—19世纪确立了板的弯曲理论；英国弹性力学家拉弗(Love，A.E.H.)建立了弹性薄壳理论；法国

力学家、几何学家圣韦南(Saint-Venant,A. J. C. B. de)研究了弹性柱的弯曲和扭转问题.在弹性力学发展的早期,大都研究线性问题,即在各种方程和边界条件中只近似到线性关系.弹性力学的发展晚期则考虑非线性项带来的修正,并逐步发展为非线性弹性力学.

塑性力学(plasticity) 固体力学的一个分支.是研究固体在外力和温度等因素作用下产生塑性变形的学科.塑性变形是指当外力或外部其他作用取消后,物体不能恢复原状的变形,亦称永久变形.物体产生塑性变形时,物体内产生塑性变形区,简称塑性区.在塑性区,应力应变关系不再是弹性关系了,而必须考虑塑性的本构关系.塑性本构关系的复杂性在于所产生的应变不仅与当前的应力张量有关,还与这一点的应力状态的历史过程有关.迄今关于它的研究还在进行,并且各学派之间还没有统一的认识.但目前大多数工程中采用的理论是塑性增量理论和塑性全量理论.塑性增量理论认为:存在一个判定材料有无塑性变形的条件,称为屈服条件.塑性本构关系本质上是增量关系,且一般可以表示为$f(\sigma_{ij})=c$,f 称为屈服函数,σ_{ij}为应力分量,c 为常数.当应力张量满足屈服条件时,该点即产生塑性应变.塑性全量理论认为:塑性本构关系是应力和应变之间的全量关系,在不缺载的条件下,与非线性弹性力学的应力应变关系类似.它的计算公式较简单,但不能反映复杂的加载历史.塑性力学问题的复杂性还在于整个问题中不同加载过程的弹塑性区的交界面是一个未定的边界,增加了一个活动边界,这给在数学上整体求解造成了一定的困难.塑性力学所研究的实际问题很多,可以说凡是弹性力学研究的课题,只要以塑性本构关系取代胡克定律,都可以作为塑性力学的课题,所以在塑性力学中也有相应的梁、板、壳理论和相应的平面问题、扭转问题等.

流体-固体耦合问题(fluid-solid coupling problem) 力学中的一类重要问题.是流体同固体相互作用,需将二者联立求解的问题.如考虑水坝振动时库容水对坝体的压力,考虑机翼变形时空气对机翼的升力,考虑贮油罐中液体晃动和罐的变形时罐底的提离问题等.这类问题的求解需要将流体力学和固体力学的全部方程联立起来才能求解,因此它是流体力学和固体力学的交叉学科.由于问题的复杂性,在大部分情况下需要作一定的简化假定才能求解.它在海洋工程、核工程、航空、航天、水利、化工、机械、土木、冶金、医学工程等领域中都经常遇到,并且有时是关键性的技术问题.

计算力学(computational mechanics) 力学的分支学科.是借助计算机探索力学规律、求解力学问题、加工力学数据的学科.20 世纪 60 年代形成独立学科.其主要内容为:

1.力学问题中的数值方法.涉及连续介质力学方程组的离散化方法,如有限单元法、加权余量法、差分法、边界积分法等;还涉及力学问题常见数学课题的数值方法,如稀疏矩阵的有关算法、样条插值、非线性问题大范围求解及分叉问题的数值方法等.

2.计算力学软件.研究和发展解决力学问题的专用或通用程序及软件系统.它又是软件工程的一个专门的分支.

3.力学数据的计算机处理.研究和发展从大量力学数据中提取必要的信息、力学量,及其有关的几何量的自动生成、图形显示的方法,以及大量力学数据的压缩存储.

目前,计算力学已在结构分析、流体力学、多刚体动力学等方向上取得重要进展.近代土木工程、航空航天工程、机械、化工各种复杂系统有关的力学问题大多诉诸于计算力学手段来解决,计算力学已成为计算机辅助设计(CAD)的组成部分.计算力学是力学、数学和计算机科学的交叉学科,它的研究成果不仅影响了力学学科,而且转移到其他学科,如计算物理、数值天气预报、数值工程地质等.

理论力学(theoretical mechanics) 力学的一个分支学科.是用推理的或数学的方法研究各种力学现象的力学领域,而不同于应用力学(参见"应用力学").理论力学也是理工科专业的一门技术基础课程,由静力学、运动学和动力学三部分组成.静力学研究力学的简化和平衡,运动学从几何学观点研究物体的运动,动力学则研究物体的运动与作用于物体的力之间的关系.理论力学课程主要讲述牛顿力学的基本原理,也包含一部分分析力学内容(参见"牛顿力学"与"分析力学").

牛顿力学(Newtonian mechanics) 力学中最古老的部分,是一切力学分支学科的基础.指在牛顿定律基础上建立起来的力学理论.其研究对象为由大量分子组成、且运动速度远小于光速的宏观物体.在人类长期实践和历代学者研究的基础上,1687 年,英国数学家、物理学家牛顿(Newton,I.)在《自然哲学的数学原理》一书中提出了质量、时间、空间和力等基本概念,总结出制约物质宏观机械运动的基本规律,即万有引力定律和动力学基本定律,奠定了牛顿力学的基础.1670 年,牛顿和德国数学家、哲学家莱布尼茨(Leibniz,G. W.)创立了研究力学规律的数学方法——微积分,从而使力学有可能从归纳性科学转变为演绎性科学,即以牛顿力学为基本出发点,利用数学推理得出结论,用以解释或预测实际力学现象,并接受实践的检验.

分析力学(analytical mechanics) 力学的一个分支.是用纯粹分析方法处理受约束力学系统的力

学分支.其特点是用对广义坐标、能量和功的分析代替牛顿力学对力和加速度的分析.分析力学与牛顿力学是经典力学的两个组成部分.1788年,法国数学家、力学家拉格朗日(Lagrange,J.-L.)提出以虚位移原理和达朗贝尔原理作为力学的演绎基础,建立了受约束力学系统的动力学普遍方程,以此为基础演化出各种形式的动力学方程,如拉格朗日方程、正则方程、阿佩尔方程等.力学的变分原理是分析力学的另一组成部分,它将真实发生的运动与可能发生的运动加以比较,并提供能甄别出真实运动的准则,以哈密顿原理和高斯原理为代表.由于计算机的发展,分析力学已成为对复杂工程对象作动力学计算的有力工具.由于分析力学的原理也适用于离散系统以外的更广泛领域,它也是从经典力学向现代物理学过渡的桥梁.

应用力学(applied mechanics) 力学的一个分支学科.是侧重于工程应用为目的的研究课题和研究方法的力学领域,如航空力学、建筑力学、船舶力学、车辆力学、陀螺力学等.它是相对于理论力学而言的(参见"理论力学").

断裂力学(fracture mechanics) 亦称裂纹力学.固体力学的一个分支.它研究含裂纹物体(简称裂纹体)的强度和裂纹扩展规律,根据裂纹尖端附近材料塑性区的大小,可分为线性弹性断裂力学和弹塑性断裂力学;根据研究的方法,可分为理论(或分析)断裂力学和统计(或概率)断裂力学;根据引起材料断裂的荷载性质,可分为断裂静力学和断裂动力学.断裂力学萌芽于20世纪20年代,格里菲思(Griffiths,A.A.)对玻璃低应力脆断的研究.其后,国际上一系列重大低应力脆断灾难事故的出现,促进了该学科的研究和发展,20世纪50年代这门学科开始形成.它的任务是:求得各类材料的断裂韧度;确定裂纹体在外荷载作用下和作用的过程中是否发生断裂和裂纹扩展规律,建立断裂准则;研究裂纹体在腐蚀环境和应力共同作用下的断裂问题.近年来在断裂力学的计算方面,有限单元法、差分法、边界元法、杂交单元法和复变函数法等已被广泛采用.目前,该学科已在航天、航空、交通运输、化工、机械、能源、材料等工程中得到广泛应用.

结构力学(structural mechanics) 固体力学的一个分支.狭义的结构往往指的是杆件结构,通常所说结构力学是指杆件结构力学,是研究杆件结构受力、传力规律以及结构创新的学科.其任务是:研究结构在荷载、温度变化和支座移动等因素作用下的强度、刚度和稳定性的计算原理和计算方法,以及结构的组成规律、合理形式和结构的力学性能.结构力学不是纯粹计算的学科,实验对结构力学的发展有着重要的意义,特别是新型或复杂的结构尤其重要,因此结构力学必须充分利用实验结果.在实验基础上,分清主要和次要因素,做出较切合实际的假定,略去次要因素,使复杂结构得以简化,建立既能充分反映结构真实受力性能和几何特性、又便于计算的力学模型,并对它进行计算,确定结构的内力和变形,由此判断结构的强度、刚度和承载能力是否满足设计要求,必要时,还应对整体或局部结构进行强度、刚度和稳定性实验.根据其研究的性质,可分为结构静力学、结构动力学、结构稳定性理论、结构的断裂和疲劳分析;根据其研究方式,可分为分析结构力学、计算结构力学和实验结构力学.

多体系统动力学(dynamics of multi-body system) 力学的一个分支.是研究多个刚体或变形体之间相互联系组成的系统的动力学行为的力学分支.多个刚体通过铰链约束相互联系构成的系统称为多刚体系统.若考虑系统内各分体的变形,则称为多柔体系统,或简称多体系统.虽然经典力学方法原则上可用于建立任意系统的微分方程,但随着系统内分体数和自由度的增多,以及分体之间约束方式的复杂化,方程的推导过程变得极其繁琐.为适应现代计算技术的飞速发展,要求将传统的经典力学方法针对多体系统的特点加以发展和补充,从而形成多体系统动力学的新分支.为建立多体系统动力学的数学模型,已经发展了各种方法,其共同特点是将经典力学原理与现代计算技术结合,以形成面向计算机的、程式化的和高效率的计算方法,这些方法可归纳为两类,即相对坐标方法和绝对坐标方法.

最早的相对坐标方法为罗伯森-维滕堡方法,其特点是:引入图论工具描述多体系统内各分体联系状况的拓扑结构,以分体间的相对位移为广义坐标,利用分析力学中的动力学普遍方程,导出多体系统动力学方程.根据拓扑结构的不同特点,多体系统可分为树系统和非树系统.对非树系统,必须用铰切剖方法转变成树系统处理.凯恩方法以广义速率代替广义坐标描述系统的运动,并将矢量形式的力和惯性力向特定的基矢投影,以消除理想约束力,因而兼有矢量力学和分析力学的特点.变分方法从分析力学中的高斯原理出发,用求极值的优化算法直接计算系统的运动规律,主要应用于工业机器人.

绝对坐标方法以各分体质心的笛卡儿坐标和描述各分体相对质心姿态的欧拉参数作为多体系统的绝对坐标.由于各分体之间存在约束,绝对坐标为非独立变量.利用分析力学中的拉格朗日乘子方法建立动力学方程,与约束方程联立求解.虽然绝对坐标方法的未知变量数远大于相对坐标方法,但由于计算过程极为程式化,且动力学方程不受系统拓扑结构变化的影响,因此也在工程计算中实际应用.

由于多体系统动力学的主要研究对象为大量分

体组成的大型复杂系统，因此不论上述哪种方法都必须利用电子计算机作数值仿真计算. 国际上已有不少商业化的通用软件，国内也已开发出有关的软件，并已付诸于使用. 对于多柔体系统，还必须与连续介质力学中的离散化方法，如有限单元法结合进行计算.

实验力学(experimental mechanics)　力学的一个分支. 是与新技术紧密结合、以实验为手段进行力学研究的一个学科. 实验力学在有关力学的基础研究、应用基础研究和工程实际力学问题的研究中，通过实验，去取得认识客观世界的第一手资料，测量如速度、变形、强度等一系列的力学量，建立能反映力学本质的模型，探索外界因素和内在力学性质之间的规律，或对工程结构的可靠性和安全性做出判断. 在人类历史上，为了生产的需要，早就有了力学实验，但作为近代科学的力学实验，直接地与力学理论基础建立密切联系，可从 17 世纪算起. 随着新技术的迅速发展，实验力学的面貌日新月异. 实验力学可以从实验原理的角度做如下分类：

1. 电学方法. 电阻应变计是典型的传感器. 各种传感器将力学量转化为电阻、电容、电感、电压等电学量，而且这些传感器具有力学量与电学量之间确定的和稳定的关系. 电学方法对实时测量和长期监控特别便利.

2. 光学方法. 基于应力、光律的光弹性方法已在工程中被广泛采用. 云纹法也有充分的发展和应用. 自从有了激光器，发展了全息干涉法、散斑干涉法和云纹干涉法，这些方法均以很高的灵敏度获得位移或振幅的信息. 焦散线法以直接获得裂端应力强度因子的特性，在断裂力学中发挥了很大的作用.

3. 声学方法. 超声扫描、超声显微镜和声全息，可以以很高的分辨率，测量物体内部的信息，且均以成像为目的. 另一类型的声——超声技术，是以强度、韧度等主要力学参量的微缺陷总体和物性异常为主要检测对象，如声弹性(测量残余应力)和声发射技术(测量含裂纹试件的开裂或监控疲劳裂纹扩展).

4. 磁学方法. 巴克豪生效应可以进行应力测量、评估缺陷和判断晶粒大小. 磁致伸缩及其逆效应可以测量残余应力. 磁声发射对材料微结构和应力均敏感. 漏磁通已从磁粉探伤发展出各类磁敏感器件，配合微机数据采集、分析和处理，可以确定缺陷的位置和深度.

5. 射线方法. X 射线透视、软 X 射线扫描显微镜、软 X 射线全息和层析(CT)技术等，它们的共同特性是利用其穿透性检测内部的状态；其波长短而空间分辨率高，可以检测在外力作用下微小的缺陷或微裂纹的演化. 其他射线，如 γ 射线或同步辐射光，均是可以利用的射线源.

6. 热学方法. 有被动式和主动式. 被动式有外部热源，是指从传播异常来发现缺陷；主动式是指由裂纹面上的摩擦发热来发现缺陷. 红外热像仪用伪彩色显示热图；热敏液晶提供真彩色热图. 复合材料在层面中的热传导，比沿厚度方向的热传导约高一个量级，热学方法适合于检测复合材料的局部损伤.

7. 其他技术或综合技术. 如脆漆法、与计算结合的混合法、光声和光热技术或磁声发射技术等，多彩纷呈，必须针对具体问题来选择合适的技术. 实验力学从获得信息的类型来分类，可以分为宏观法(如焦散线法、声发射、磁声发射等)、逐步法(如电测法、磁致伸缩以及某些光学方法的逐点分析)和全场法(各种声、光、热、射线等的成像方法). 实验力学从测试对象来分类，可分为模拟法和实测法. 实测法是直接对研究对象进行测量. 模拟法包括基于同类物理方程的比拟法(如用薄膜在气压下的高度分布，比拟柱体的扭转应力的薄膜)和基于相似律的模型法(如大型水坝的缩尺模型实验等).

实验力学的发展与新技术的发展密切相关. 如激光器的出现，计算机的普及，多点数据采集和处理能力的提高，图象处理系统功能的迅速加强，层析技术的出现等，都不断地把实验力学推向一个新的高度，在认识世界和改造世界方面，发挥愈来愈大的作用，成为科学研究及航空、航天、造船、石油、机械、水利、建筑等各种工业建设中不可缺少的重要方面.

本构关系(constitutive relationship)　应力张量与应变张量的关系. 一般地，指将描述连续介质变形的参量与描述内力的参量联系起来的一组关系式. 具体地讲，指将变形的应变张量与应力张量联系起来的一组关系式，又称本构方程. 对于不同的物质，在不同的变形条件下有不同的本构关系，也称为不同的本构模型. 如线性弹性体本构模型的本构关系为胡克定律. 弹塑性体、粘弹性体、热弹性体等都有各自的本构关系. 刚体也可以被看做一种简单的本构模型.

离散化方法(discretization method)　一种求解连续介质力学问题的基本手段. 是用有限个参量来近似表征连续介质力学中物理量的方法. 它是用计算机求解连续介质力学问题的. 连续介质力学规律一般用微分方程或积分方程组来描述，离散化方法将它们化归为有限个参量的代数方程组来近似原问题. 通常的有限差分法、有限单元法、加权余量法都是行之有效的离散化方法.

相似律(similarity law)　力学的基本规律之一. 指不同的物理或力学现象之间相似的规律. 这里的相似，一般理解为从一个现象有关的物理量，经过适当的变换可以得到另一现象相应的物理量. 通常

有三种类型的相似：

1. 模型相似. 用模型试验取得的有关物理量来推算原型相应的物理量，这时模型与原型的已知几何大小与物理量间必须满足模型相似律，也称模型律.

2. 单因素相似. 它是模型相似的某种近似. 在物理现象的众多因素中，要求各种因素都满足相似条件通常是不可能的，这时经常舍弃一些次要的条件，只要求主要条件满足，如在流体力学风洞模型试验中只要求雷诺数相等.

3. 数学比拟. 原则上不同的两类物理现象，由于它们满足的控制方程相同，所以相应的物理量之间有比拟关系. 如薄膜在均匀压力下平衡和弹性柱体扭转问题，前者的挠度和后者的应力函数都满足泊松方程，利用这种相似建立起来的用薄膜求解扭转问题的装置称薄膜比拟实验.

相似律是用实验方法求解力学问题的理论基础.

量纲分析(dimensional analysis)　一种对不同物理量之间关系的分析方法. 当基本物理量的单位尺度改变时，其他物理量随之变化的规律的分析方法. 基本物理量为长度、质量和时间，它们的单位若表示为 L，M，T，则当单位改变时，这些量是线性变化的. 于是将长度、质量、时间的量纲定义为 L，M，T. 其他物理量当基本量尺度改变时，一般改变为 L^{α}，M^{β}，T^{γ}，是 L，M，T 的非线性规律变化，其量纲记为 L^{α}，M^{β}，T^{γ}. 这些物理量也称为导出量. 常见力学量的量纲如下表：

物　理　量	量　　纲
长度	L
质量	M
时间	T
速度	LT^{-1}
加速度	LT^{-2}
密度	ML^{-3}
力	MLT^{-2}
功或能	$ML^{2}T^{-2}$
应力或压强	$ML^{-1}T^{-2}$

量纲分析还给出各物理量处于一个方程中时，方程的各项应具有相同的量纲的结论. 这一结论称为量纲的齐次性原则. 改变度量单位的尺度，在一定程度上讲，和在同一单位尺度下观察两个不同规模的物理过程是同一件事，而后者是模型律所讨论的内容. 所以量纲分析与相似律有着密切的关系.

波(wave)　一种力学现象. 是连续介质中扰动随时间传播的现象. 在波传播过程中，介质中的质点做振动或周期运动. 波的种类很多：小扰动在空气中传播称为声波；小扰动在弹性介质中传播称为弹性波；在传播过程中伴有塑性变形时，产生塑性波；质点运动方向与波传播方向相同时称为纵波；与波传播方向垂直时称为横波；波沿自由表面传播时称为表面波，如水波；弹性介质的表面波称为瑞利波；波沿两种介质的交界面传播时称为界面波；波传播过程中伴有应力间断时称为激波或冲激波. 波在传播中遇到表面或界面时，会产生反射或折射现象. 通常又将固体中的波称为应力波；描述波动的方程为线性方程时称为线性波；否则称为非线性波. 波与工程技术关系密切，爆破工程、抗震工程、航空、航天、防护等工程领域都需要对有关的波传播与检测有深入的了解. 在研究连续介质中波的现象时，一般需要在连续介质力学有关的方程组中考虑由于质量分布引起的惯性力.

自治系统(autonomous system)　亦称自守系统. 一种特殊的动力系统. 指用一种特别类型微分方程组描述的有限自由度运动. 有时也将这种微分方程组称为自治系统. 令 $x \in \mathrm{R}^n$ 为相空间中的点，则自治系统的形式为

$$\dot{x} = F(x), \quad F \in \mathrm{R}^n,$$

右端函数 F 不显含时间 t. 有时需要讨论含参数的自治系统，令 $\lambda \in \mathrm{R}^m$，含参数自治系统的形式为

$$\dot{x} = F(x,\lambda), \quad x \in \mathrm{R}^n, \lambda \in \mathrm{R}^m, F \in \mathrm{R}^n.$$

通常多讨论 $m=1$，即含单个参数的自治系统. 如果系统右端项相依于时间 t，则一般称这种系统为非自治系统. 自治系统和非自治系统有时统称为动力系统. 动力系统有时也称为微分动力系统，一般可写为

$$\dot{x} = F(t, x, \lambda), \quad x \in \mathrm{R}^n, \lambda \in \mathrm{R}^m, F \in \mathrm{R}^n.$$

当 n 有限时，称为有限维动力系统，这时它是常微分方程组；当 n 无限时，称为无限维动力系统，这时 F 为定义在函数 x 上的算子，它是偏微分方程或方程组. 力学中研究的对象一般都可以用动力系统来描述，物理、化学、经济学等领域的问题有时也可以用动力系统来描述.

非自治系统(nonautonomous system)　见“自治系统”.

微分动力系统(differential dynamical system)　见“自治系统”.

定常解(steady solution)　微分动力系统的特殊解. 令动力系统为

$$\dot{x} = F(t,x,\lambda), \quad x \in \mathrm{R}^n, \lambda \in \mathrm{R}^m, F \in \mathrm{R}^n,$$

则定常解满足 $\dot{x}=0$，即 $F(t, x, \lambda)=0$. 通常在自治系统情形，定常解为 $x(\lambda)$，满足 $F(x(\lambda),\lambda)=0$，是 (x,λ)空间中的一维流形. 定常解有时亦称平衡解.

平衡解(equilibrium solution)　见“定常解”.

李亚谱诺夫稳定性理论(Liapunov theory of stability)　判断运动是否稳定的理论.由俄国数学家、力学家李亚普诺夫(Ляпунов, A. M.)于 1892 年奠定基础,其后发展成为关于运动稳定、渐近稳定、不稳定的定义和判别方法的理论体系.李亚普诺夫关于运动稳定性的定义是:令动力系统为 $\dot{x}=F(t,x)$, $F,x\in \mathrm{R}^n$,当 $t=t_0$ 时,初始条件为 x_0,有解 $x^*(t)=x^*(t,t_0,x_0)$.设对于扰动的初始条件 $t=t_0$ 时,$x(t_0)=x_1$,解为 $x(t)=x(t,t_0,x_1)$,则对于任意 $\varepsilon>0$,存在 $\delta>0$,当 x_1 满足 $\| x_0-x_1 \| <\delta$ 时,永远有

$$\| x^*(t)-x(t) \| <\varepsilon,$$

则称 $x^*(t)$为稳定的.若 $x^*(t)$为稳定的,且当 $t\to\infty$ 时有

$$\| x^*(t)-x(t) \| \to 0,$$

则称 $x^*(t)$为渐近稳定的.若 $x^*(t)$不满足稳定条件,则称它为不稳定的.如果 $x^*(t)$是平衡解,则相应地得到平衡稳定、渐近稳定和不稳定.李亚普诺夫还给出了判定系统稳定性的方法.

莫夫强运动稳定性定义(Movchan definition of stability of motion)　对于李亚普诺夫运动稳定性定义的一种推广.是由苏联学者莫夫强(Мовчан, A. A.)于 1960 年提出的.其内容是:考虑动力系统 $\dot{x}=F(t,x)$在 $t=t_0$ 时初条件 x_0 的解 $x^*(t,t_0,x_0)$和 $t=t_0$ 时初条件 x_1 的解 $x(t,t_0,x_1)$,并定义两个半模 $\rho_1(x_0,x_1)$和$\rho_2(x^*,x)$.如果对于任意 $\varepsilon>0$,都存在 $\delta>0$,而且当 $\rho_1(x_0,x_1)<\delta$ 时,永远有

$$\rho_2(x^*,x)<\varepsilon,$$

则称解 x^* 是在莫夫强意义下稳定的;在稳定条件下,当 $t\to\infty$时,有 $\rho_2(x^*,x)\to 0$,称为在莫夫强意义下渐近稳定的;不满足稳定性条件称为不稳定的.莫夫强稳定性定义有明显的优点:第一,它适用于无限自由度系统,连续介质系统(如流体或弹性系统)的稳定性问题可以纳入这个定义来讨论.第二,它引进半模 ρ_1,ρ_2,脱离通常的 n 维空间距离,可以对于特定的物理量来讨论稳定性问题.

平衡解的分岔(bifurcation of equilibrium solution)　力学系统的基本概念之一.指含参数 λ 的动力系统在同一 λ 邻近,平衡解 $x(\lambda)$具有多解的现象.动力系统

$$\dot{x}=F(x,\lambda),\quad F,x\in \mathrm{R}^n,\ \lambda\in \mathrm{R}$$

的平衡解(x^*,λ^*)满足 $F(x^*,\lambda^*)=0$,如果 $\lambda^*+\delta\lambda=\lambda$ 有两个不同的 x_1,x_2,使

$$(x_1,\lambda^*+\delta\lambda),\quad (x_2,\lambda^*+\delta\lambda)$$

皆为平衡解,则称(x^*,λ^*)为上述含参数自治系统平衡解的分岔点.具有分岔点的平衡解流形$x(\lambda)$称为分岔解.分岔点常常是稳定平衡点和不稳定平衡点在平衡解曲线上的分界点.如果从某一平衡点(x_0,λ_0)出发,沿平衡解曲线前进,首次遇到的分岔点称为首次分岔点.从分岔点沿平衡解曲线一支前进再遇到新的分岔点称为二次分岔点,也称二次分岔.复杂系统会有多次分岔.分岔点常常是力学上平衡的临界状态.确定分岔点,无论在实际工程上,还是在理论上都具有重要意义.工程中临界荷载、临界速度都对应于分岔点的参数值.分岔点所必须满足的条件是

$$\mathrm{rank}\left(\frac{\mathrm{D}F}{\mathrm{D}x}\right)<n \quad 或 \quad \left\| \frac{\mathrm{D}F}{\mathrm{D}x} \right\| =0,$$

这里 rank 为秩,$\frac{\mathrm{D}F}{\mathrm{D}x}$为 F 的梯度矩阵.

分岔点(bifurcation point)　见“平衡解的分岔”.

分岔解(bifurcation solution)　见“平衡解的分岔”.

霍普夫分岔(Hopf bifurcation)　一种分岔现象.是带参数的自治系统平衡解转变为振动的分岔现象.该现象由苏联物理学家、数学家安德罗诺夫(Андронов, A. A.)于 1930 年和德国数学家霍普夫(Hopf, E.)于 1942 年发现.令含参数的自治系统为

$$\dot{x}=F(x,\lambda),\quad F,x\in \mathrm{R}^n,\ \lambda\in \mathrm{R},$$

并令(x^*,λ^*)为平衡解,考虑 $F(x,\lambda)$在(x^*,λ^*)邻近展开,有

$$F(x,\lambda)\approx \frac{\mathrm{D}F}{\mathrm{D}x}(x-x^*).$$

引进新变量 $y=x-x^*$,有

$$\dot{y}=\frac{\mathrm{D}F}{\mathrm{D}x}y.$$

当 $A=\mathrm{D}F/\mathrm{D}x$ 的全部特征值 $\mu_1,\mu_2,\cdots,\mu_n$ 具有负实部时,平衡点为稳定的.若 $A=\mathrm{D}F/\mathrm{D}x$ 的某对特征值的实部为 0 时,即当有一对纯虚的特征值时,对应于霍普夫分岔点.在霍普夫分岔点邻近,系统会存在稳定或不稳定的周期运动,这种运动称为极限环.霍普夫分岔现象在力学中很普遍.对含参数的非线性振动系统,当参数达到某值时产生的振荡,称为参数振荡.飞行时当速度达到一定值时会产生机翼的颤振,这些都与霍普夫分岔有关.20 世纪 40 年代,霍普夫和苏联学者朗道(Ландау, Л.)提出猜想,认为湍流是由于多次霍普夫分岔所形成,这就是著名的霍普夫-朗道猜想.

极限环(limit cycle)　动力系统的孤立周期解.非线性自治系统孤立的周期特解.它在相空间表现为孤立的闭轨线.只有在非保守力作用下的非线性自治系统才有孤立的(即其无限邻近不存在别的)周期解.极限环有稳定的和不稳定的两类,前者运动受

扰动后，当$t\to\infty$时趋于极限环，而后者当受扰动后远离极限环.

平衡稳定性的能量判据(energy criterion of equilibrium stability)　判定力学系统平衡状态是否稳定的一种准则. 这个判据最早的提法是意大利数学家、物理学家托里切利(Torricelli，E.)于1644年提出的准则：重力作用下的物体，当重心取最低时保持平衡且是稳定的. 其后，这一准则被法国数学家、力学家拉格朗日(Lagrange，J.-L.)于1788年在他的名著《分析力学》中总结为一般的原理，亦称拉格朗日原理，即当保守系统处于势能孤立极小状态时，系统处于稳定平衡. 令系统为离散系统，广义坐标为$x\in\mathrm{R}^n$，并令x为平衡状态，系统的势能为$V(x)$，则当对于x的任意小偏移δx，均有

$$\delta^2 V=\frac{\partial^2 V}{\partial x_i \partial x_j}\delta x_i \delta x_j>0$$

时，系统在x的平衡为稳定. 如果存在某个小偏移δx^*，使

$$\delta^2 V=\frac{\partial^2 V}{\partial x_i \partial x_j}\delta x_i^* \delta x_j^*<0,$$

则称系统在x的平衡是不稳定的. 上面的条件如果变为

$$\delta^2 V=\frac{\partial^2 V}{\partial x_i \partial x_j}\delta x_i^* \delta x_j^*=0,$$

称系统沿方向δx^*为随遇平衡. 对于弹性系统的势能需计入变形能.

平衡稳定性的静力学判据(statics criterion of equilibrium stability)　判定力学系统平衡状态是否稳定的一种准则. 内容是：令在离散的力学系统上取广义坐标$x\in\mathrm{R}^n$，并令x为系统的平衡位置，$Q\in\mathrm{R}^n$为广义力向量，则在平衡位置上$Q_i(x)=0(i=1,2,\cdots,n)$. 如果系统偏离平衡位置任意一个小增量δx，同时广义力获得一个增量为δQ，而且有

$$\delta Q\cdot\delta x<0,$$

则称系统在x处的平衡为稳定的. 如果存在平衡位置的小偏移δx^*，使

$$\delta Q\cdot\delta x^*>0,$$

则称系统是不稳定的. 如果上面条件中变为

$$\delta Q\cdot\delta x^*=0,$$

则称平衡是临界的，或沿δx^*是随遇平衡. 对于弹性系统来说，在广义力中需增加变形产生的恢复力. 静力学判据是一种不十分严格的、但简便易行的判据，对于系统有势时，它符合李亚普诺夫稳定性判据，对于非保守力则往往有例外的反例.

力学单位制(system of mechanical units)　力学量的单位体系. 通常有两种：

1. 绝对单位制，以长度、质量、时间为基本单位，其他为导出单位. 其中以长度单位为米(m)、质量单位为千克(kg)、时间单位为秒(s)作为基本单位的，称为米千克秒制(MKS制)，这是国际单位制(代号SI)的力学单位. 例如，使1千克质量的质点产生1米·秒$^{-2}$加速度的作用力为1牛顿(N)，这里牛顿是力的单位，是导出单位. 而将长度为厘米(cm)、质量为克(g)、时间为秒(s)作为基本单位的，称为厘米克秒制(CGS制). 例如，使1克质量的质点产生1厘米·秒$^{-2}$加速度的作用力为1达因(dyn)，1dyn=10^{-5}N，这里达因是力的单位，是导出单位.

2. 重力单位制，以长度、力、时间为基本单位，其他量为导出单位. 例如，在这里质量成为导出单位. 而力的单位是相当质量的物体在重力加速度为$G=9.80655$米·秒$^{-2}$(实用上常近似地取$G=9.8$米·秒$^{-2}$)的地点所受的力. 例如，1克质量的物体受的力称为1克力，1千克质量的物体受的力为1千克力，1磅质量的物体受的力为1磅力. 其中以长度为米(m)、力为千克力(kgf)、时间为秒(s)作为基本单位的，是常用的工程单位制. 1工程单位的质量(EUM)是在1千克力作用下产生1米·秒$^{-2}$加速度的质点所具有的质量. 也有取长度为厘米(cm)、力为克力(gf)、时间为秒(s)作基本单位的，但不常用. 若取长度为英尺(ft)、力为磅力(lbf)、时间为秒(s)作基本单位，则称为英工程制. 1英制工程单位的质量称为斯勒格(slng)，是在1磅力作用下产生1英尺·秒$^{-2}$加速度的质点所具有的质量.

几种力学单位制

力学量	SI	CGS制	工程单位制	英工程制
长度	米	厘米	米	英尺
质量	千克	克	千克	斯勒格
力	牛顿	达因	千克力	磅力
功	焦耳	尔格	千克力·米	英尺·磅力
功率	瓦特，千瓦	尔格·秒$^{-1}$	千克力·米·秒$^{-1}$，	磅力·英尺·秒$^{-1}$

帕斯卡(pascal)　简称帕. 力学中量的单位. 压力强度的单位，用符号Pa表示.

$$1\mathrm{Pa}=1\mathrm{N}\cdot\mathrm{m}^{-2}=1\mathrm{m}^{-1}\cdot\mathrm{kg}\cdot\mathrm{s}^{-2}.$$

例如，1标准大气压=1.013×10^5Pa；1工程大气压=1kgf·cm^{-2}=0.98×10^5Pa；1巴(用符号bar表示)=10^5Pa，而1毫巴(mbar)=10^{-3}巴，是气象方面常用的大气压单位. 法国数学家帕斯卡(Pascal，B.)于1653年提出了著名的帕斯卡定律：对密闭容器中的流体某处施加压力，可传到流体各处而不减小压力强度. 压力强度以帕为单位是后人对帕斯卡的纪念.

毫巴(millibar)　力学中量的单位. 是气象领域

常采用的大气压力单位，常用符号 mbar 表示. 地面上的大气压力约为 1000mbar＝1bar＝10^5Pa.

大气压(atmosphere) 力学中量的单位. 指地面上的空气压力强度. 1 标准大气压为 76cm 汞柱高，等于 1.013×10^5Pa，工程大气压为

$$1\text{kgf}\cdot\text{cm}^{-2}=0.98\times10^5\text{Pa},$$

都近似等于

$$1000\text{mbar}=1\text{bar}=10^5\text{Pa}.$$

达因(dyn) 见“力学单位制”.

瓦特(watt) 简称瓦. 力学中量的单位. 指国际单位制中功率的单位，用符号 W 表示.

$$1\ \text{W}=1\ \text{J/s}=1\ \text{N}\cdot\text{m/s}=1\ \text{kg}\cdot\text{m}^2/\text{s}^3,$$

即每秒作功 1 焦耳. 工程上常用的功率单位是千瓦，记为千瓦或 kW，等于 1000 W；或者是公制马力，记为 ps，近似取为 735.5W.

尔格(erg) 力学中量的单位. 指厘米克秒(CGS)制中功的单位，用符号 erg 表示，

$$1\ \text{erg}=1\text{dyn}\cdot\text{cm}=1\text{g}\cdot\text{cm}^2\cdot\text{s}^{-2},$$

即 1dyn 力沿其方向移动 1cm 所作的功.

$$1\text{J}=10^7\text{erg}.$$

马力(horse power) 力学中量的单位. 指工程上常用的功率单位. 有英制马力(hs)和公制马力(ps)两种，通常是指公制马力 ps，$1\text{ps}=75\text{kgf}\cdot\text{m}\cdot\text{s}^{-1}$(参见“力学单位制”).

焦耳(joule) 力学中量的单位. 指国际单位制中功的单位，简称焦(J).

$$1\ \text{J}=1\ \text{N}\times1\ \text{m}=1\ \text{N}\cdot\text{m}=1\ \text{kg}\cdot\text{m}^2\cdot\text{s}^{-2},$$

即 1N 的力沿其方向移动 1m 所作的功.

理 论 力 学

静力学(statics) 力学研究的重要分支. 是研究物体在力系作用下平衡规律的学科. 力学的各分支学科都存在静力学问题，如流体的静力学、弹性体的静力学、塑性体的静力学等. 在理论力学中静力学主要针对刚体. 它讨论作用于刚体的力系等效替换、简化、平衡条件和应用. 因为对刚体平衡的必要且充分条件对变形体只是必要条件，故理论力学中得到的静力学平衡条件可按必要条件用于变形体力学，例如弹性力学、结构力学、流体力学的平衡方程就是这样建立的. 在工程技术中刚体的静力学理论可直接应用于计算和设计.

力(force) 力学的基本概念之一. 是物体之间的相互作用，是物体机械运动状态发生变化或变形的原因. 其大小、方向、作用点决定力对物体作用的效应，称为力的三要素. 可通过作用效应确定力的大小和方向，而作用点是力作用位置的抽象. 对物体相互接触而产生的力，当接触面积相对很小时，可近似认为作用于一点，称为集中力；接触面积不太小时称为面分布力，如风对墙壁和水对坝体的压力. 分布力中还有分布于物体内每一点的体分布力，如重力.

从宏观上看，力可分为两类：一是物体相互接触而产生的，如摩擦力、拉压力；二是通过场而发生的，如万有引力、电磁力. 从微观上看，20 世纪 60 年代末，物理学归纳出宇宙有四种相互作用力，即：引力相互作用，存在于具有质量的所有物体之间；电磁力相互作用，存在于荷电的物体之间；强相互作用，存在于核子(中子和质子)、介子和超子之间；弱相互作用，是基本粒子之间的另一种相互作用. 强、弱两种相互作用是近距离的作用(只在 $10^{-17}\sim10^{-15}$m 内才起作用). 宏观物体之间，重要的是前两种力. 接触力是原子、分子、电磁相互作用的总和，场力即引力和电磁力.

力的单位：在国际单位制之下为牛顿，在工程单位制中是千克力(kgf)或吨力(tf)，

$$1\text{tf}=1000\text{kgf},\quad 1\text{kgf}=9.80665\text{N}.$$

力的三要素(three elements of force) 见“力”.

质点(particle) 力学研究对象的一种抽象模型. 是具有一定质量的几何点. 在所研究的问题中，当物体的大小和形状都可忽略时，则视为质点，如研究天体的运行轨道、研究物体的平动等，都可把研究对象视为质点，这样可以准确地把握问题的实质.

质点系(particles system) 力学的基本概念之一. 是有限或无限个质点的集合，简称质系. 可按研究问题的需要把某些物体看做是质点系，系内各质点的联系可有多种形式，例如，刚体就是任意两质点距离不变的质点系. 一个物体或多个物体都是质点系.

刚体(rigid body) 一种特殊的质点系(其中任意两质点距离不变)，理想化的力学模型. 是不变形的物体. 力学研究的很多问题中，若略去物体的变形，则可大大减少问题的复杂程度而把握它的实质，所得的结果有的可直接用于工程实际，有的可作为进一步研究的基础. 例如，工程结构虽在受力后发生变形，但可先按刚体列出其平衡条件，求解支座反力，解决支座的设计问题；在这个基础上，再进一步研究内力分布和变形情况，解决结构本身的设计问题. 许多变形体力学都要借助刚体力学这个基础.

力系(system of forces) 力学的基本概念之一. 指一群力的集合. 其中各力在同一平面的称为平面力系，否则称为空间力系. 这两种情况，均可按力的作用线，或汇交一点，或互相平行，或不汇交又不平行，分别称为平面(或空间)的汇交力系、平行力系或任意力系.

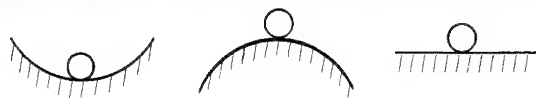
平面力系(planar force system)　见"力系".

空间力系(spatial force system)　见"力系".

平行力系(parallel force system)　见"力系".

任意力系(arbitrary force system)　见"力系".

质点系平衡(particle system in equilibrium)　静力学的基本概念之一.指质点系相对惯性系静止或作匀速直线运动.在日常生活和工程技术的多数问题中,把固定于地球的坐标系当做惯性系就足够精确了,故质点系相对地球静止或作匀速直线运动就称为平衡.对静止形式的平衡,可分为三类:受微小干扰(冲击或偏离)后,仍保持每一质点在平衡位置附近一定范围内运动的为稳定平衡;否则为不稳定平衡;平衡位置可在一定范围内连续变化的为随遇平衡.若质点系所受主动力有势,则对应这三类平



稳定平衡　　不稳定平衡　　随遇平衡

衡位置的势能依次为极小、极大和常量.简例如图.

稳定平衡(stability equilibrium)　见"质点系平衡".

不稳定平衡(non-stability equilibrium)　见"质点系平衡".

随遇平衡(indifferent equilibrium)　见"质点系平衡".

力系平衡(force system in equilibrium)　静力学的基本概念之一.指作用于质点或刚体上的力系不产生效应,即各力之效应互相抵消的情形.力系平衡的条件为:力系的主矢为零且对任一点的主矩为零.在直角坐标系下,各种力系的平衡方程可表示为:

1.空间任意力系:

$$\sum F_x = 0,\ \sum F_y = 0,\ \sum F_z = 0,$$
$$\sum m_x = 0,\ \sum m_y = 0,\ \sum m_z = 0.$$

2.空间汇交力系:

$$\sum F_x = 0,\ \sum F_y = 0,\ \sum F_z = 0.$$

3.空间平行力系(设力和 z 轴平行):

$$\sum F_z = 0,\ \sum m_x = 0,\ \sum m_y = 0.$$

4.空间力偶系:$\sum m_x=0$, $\sum m_y=0$, $\sum m_z=0$.

5.平面任意力系(z 轴垂直力系所在平面):

$$\sum F_x = 0,\ \sum F_y = 0,\ \sum m_Q = 0.$$

6.平面平行力系(令 y 轴和力平行):

$$\sum F_y = 0,\quad \sum m_Q = 0.$$

7.平面汇交力系:$\sum F_x = 0, \sum F_y = 0$.

8.平面力偶系:$\sum m_Q = 0$.

以上 F_x, F_y, F_z 分别表示力在 x, y, z 轴上的投影,m_x, m_y, m_z 分别为力对 x, y, z 轴的力矩.Q 为 xy 平面上任一点,m_Q 为力对 Q 的力矩,反时针为正,顺时针为负.除力偶系之外,各种平衡方程都可有其他形式的同个数的等价方程.例如,平面任意力系有

$$\sum m_A = 0,\ \sum m_B = 0,\ \sum m_C = 0$$

或

$$\sum m_A = 0,\ \sum m_B = 0,\ \sum F_\xi = 0,$$

式中 A, B, C 为不共线的三个点,轴线 ξ 不与 AB 直线垂直,F_ξ 为力在 ξ 上的投影.空间任意力系的平衡方程可表示为对六个轴矩等于零的形式(对六个轴有一定条件).

静力学公理(axioms of statics)　静力学理论的基础原则.它们经实践反复验证而得以公认.共有五条:

1.力的平行四边形法则.作用在物体上同一点的两个力,可合为过该点的一个力,其大小和方向由两个力向量为邻边构成的平行四边形的对角线确定.

2.二力平衡条件.刚体受两个力作用而平衡,其充分必要条件是这两个力大小相等、方向相反、作用线重合.

3.加减平衡力系原理.在已知力系上加或减去任意的平衡力系,不改变原力系对刚体的作用.

4.作用和反作用定律.作用力和反作用力总是同时存在、大小相等、方向相反,且沿同一直线分别作用于互相作用的二物体.

5.刚化原理.变形体在任意力系作用下达到平衡态,若增加约束,甚至将其约束为刚体,则平衡状态不变.

静力学的全部内容可作为以上公理的推论.如力的可传性,作用于刚体上的力可沿其作用线移到刚体上任一点而效应不变,即刚体上的力是滑动向量;又如三力平衡定理,即刚体受三力而平衡,若其中二力相交,则这三力必共面且共点,事实上,可证明刚体受任何三力平衡,该三力必共面,或汇交或平行.

二力平衡条件(equilibrium condition of two forces)　见"静力学公理".

加减平衡力系原理(principle of addition or subtraction equilibrium force system)　见"静力学公理".

作用和反作用定律(law of action and reaction)　见"静力学公理".

硬化原理(principle of solidification)　见"静力学公理".

力矩(moment of force)　力学的基本概念之

一. 是力使物体绕一点或一轴转动的状态发生改变的作用的度量. 称该点为矩心,该轴为矩轴. 力对点的矩可表示为向量,它等于由矩心到力作用点所确

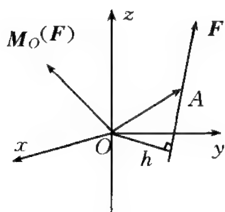


图1

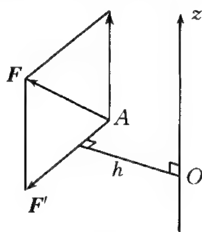


图2

定的向量和力向量的向量积. 如图1,力 $\boldsymbol{F}$ 作用点为 A,矩心点为 O,力对 O 的力矩为

$$\boldsymbol{M}_O(\boldsymbol{F}) = \overrightarrow{OA} \times \boldsymbol{F},$$

其大小为 $|\boldsymbol{F}|h$,式中 h 是 O 到 $\boldsymbol{F}$ 的距离,称为力臂. 该力矩的方向是由 $\overrightarrow{OA}$ 到 $\boldsymbol{F}$ 再到 $\boldsymbol{M}_O(\boldsymbol{F})$,符合右手螺旋法则. 力对轴的矩是代数量,把力分解为平行和垂直于轴的两个分量,垂直分量的大小乘以它与轴的距离即力矩大小,若垂直分量绕轴的转向相对轴向符合右手螺旋法则,取正号,反之取负号. 如图2,力 $\boldsymbol{F}$ 相对 z 轴的垂直分量为 $\boldsymbol{F}'$,它到 z 的距离为 h,力对 z 轴的矩为 $\boldsymbol{M}_z(\boldsymbol{F}) = |\boldsymbol{F}'|h$. 若取 $Oxyz$ 坐标系,令 $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}$ 为坐标轴上的单位向量,点 A 的坐标为 (x,y,z),设 $\boldsymbol{F}=F_x\boldsymbol{i}+F_y\boldsymbol{j}+F_z\boldsymbol{k}$,则

$$\boldsymbol{M}_O(\boldsymbol{F}) = \overrightarrow{OA} \times \boldsymbol{F} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x & y & z \\ F_x & F_y & F_z \end{vmatrix}.$$

可见,$\boldsymbol{M}_O(\boldsymbol{F})$在 z 轴的投影为

$$xF_y - yF_x,$$

正好等于 $\boldsymbol{F}$ 的垂直分量对 z 轴的力矩,亦即 $\boldsymbol{F}$ 对 z 轴的力矩$M_z(\boldsymbol{F})$. 对其他轴亦有类似情形,故力对某轴上任一点的力矩在该轴的投影,就等于力对这个轴的力矩. 力矩的单位在国际单位制中为 N·m,在工程单位制中为 kgf·m.

矩心(center of moment)　见“力矩”.

矩轴(axis of moment)　见“力矩”.

力臂(force arm)　见“力矩”.

力偶(couple)　力学的基本概念之一. 大小相等,方向相反,作用线平行而不重合的一对力称为力偶. 这二力作用线之间的距离称为偶臂. 力偶的偶矩(简称力偶矩)可以用一个向量表示,称为力偶矩向量. 该向量的模等于力偶中一力的大小乘以偶臂,方向垂直于二力的作用面,并且与二力的转向符合右手螺旋法则. 力偶对空间任一点的力矩等于其偶矩. 在刚体上,可保持力偶矩向量不变,而任意调整力的大小和偶臂长短,也可改变动力偶的作用位置. 两力偶可合为一力偶,合偶矩等于原来二力偶矩向量的合向量,即刚体上的力偶矩可看做是自由向量. 力偶矩的单位,用国际单位制是 N·m,用工程单位制是 kgf·m.

偶臂(couple arm)　见“力偶”.

偶矩(couple moment)　见“力偶”.

力的平移定理(theorem on translation of force)　作用于刚体上的力的一个等效定理. 即:作用于刚体 A 点的力,可平行移动于另一点 B,同时附加一个力偶,力偶矩等于原力对 B 点的力矩,替换后效应不变.

力的多边形法则(polygon rule of forces)　共点力系合成的一种几何方法. 将共点力系的各力矢量不计先后一个个地平行移动,使它们首尾相接构成一折线,从折线的起点到终点可连结一有向线段,该有向线段就代表合力的大小和方向. 折线在空间的起点可任意选取. 当共点力系只有两个力时,则为三角形法则,它是平形四边形法则的另一种表述.

力的平行四边形法则(parallelogram law of forces)　见“静力学公理”.

二平行力的合成(composition of two parallel forces)　关于求二力的等效力的方法. 刚体上,两个同向平行力可合成为一个力,其大小等于该二力大小之和,与二力同向,并以二力大小的反比内分二力作用点连线的点为其作用点;两个异向而大小不等的平行力可合为一力,合力的大小等于二力大小之差,与二力中较大的同向,作用点在较大一个力的外侧,并以二力大小的反比外分二力作用点的连线. 若二平行力大小相等、方向相反而不共线,则组成一力偶而无合力. 若保持二平行力的大小、作用点和相互的相对方位不变,平行地在空间旋转,则合力的大小和作用点将都不变.

平行力系的中心(centre of parallel forces)　力学的基本概念之一. 指平行力系合力的作用点. 在直角坐标系中,其坐标为

$$x_c = \frac{\sum F_i x_i}{\sum F_i},\ y_c = \frac{\sum F_i y_i}{\sum F_i},\ z_c = \frac{\sum F_i z_i}{\sum F_i},$$

式中 F_i 为各力的量值,x_i,y_i,z_i 为各力作用点坐标. 若 $\sum F_i$ 为 0,则力系或平衡,或等效于一力偶.

伐里尼翁定理(Varignon theorem)　亦称合力矩定理. 空间力系的等效简化定理,即:当力系有合力,其合力对任一点(轴)的力矩等于力系的各力对该点(轴)力矩的向量(代数)和. 该定理因法国数学家伐里尼翁(Varignon,P.)首先提出而得名.

合力矩定理(theorem on moment of resultant force)　即“伐里尼翁定理”.

力系的主向量(principal vector of force system)　亦称主矢. 力学概念. 指力系中各力的向量和. 它不是力,也不是力系的合力,但若力系存在合

力，则合力向量等于力系的主向量.

力系的主矩(principal moment of force system)　力学的基本概念之一. 指力系中各力对某一点力矩的向量和. 同一力系对不同点具有不同的主矩. 例如，力系对 A 点和 B 点的主矩 $\boldsymbol{M}_A$，$\boldsymbol{M}_B$ 有关系：

$$\boldsymbol{M}_B = \boldsymbol{M}_A + \overrightarrow{BA} \times \boldsymbol{R},$$

式中$\overrightarrow{BA}$是由 B 到 A 的向量，$\boldsymbol{R}$ 是力系的主向量.

力系的合力(resultant force of force system)　一种力. 指刚体上与整个力系等效的一个力，即能取代力系对刚体作用的那一个力. 一般地，力系可能有合力，也可能无合力. 例如，刚体上同方向的二平行力必有合力，但异方向大小相等的二平行力就组成一力偶，而无合力.

力螺旋(force screw)　亦称偶单力组. 刚体上空间任意力系最一般的化简形式. 指偶矩向量与力向量平行的一个力偶和一个力组成的力系. 当力与力偶矩向量同向时称为右螺旋；异向称为左螺旋. 例如，钻孔时，钻头对工件的作用就是力螺旋. 力螺旋的力作用线称为其中心轴. 若刚体上一力和一力偶的偶矩向量相互垂直，则可化为一个力. 若力和力偶的偶矩向量斜交，则可化为一个力螺旋，但其作用位置和力偶矩向量都要变化.

偶单力组(a couple and a perpendicular force)　即“力螺旋”.

右螺旋(right screw)　见“力螺旋”.

左螺旋(left screw)　见“力螺旋”.

中心轴(centre axis)　见“力螺旋”.

重心(centre of gravity)　力学的基本概念之一. 指物体重力的合力作用点. 它相对物体有确定的位置(不一定在物体上)，与物体的空间位置无关. 具有对称特点的匀质体的重心必在对称面或对称轴上. 在直角坐标系中，重心 c 的坐标为

$$x_c = \frac{\int_V xq\mathrm{d}V}{\int_V q\mathrm{d}V},\quad y_c = \frac{\int_V yq\mathrm{d}V}{\int_V q\mathrm{d}V},\quad z_c = \frac{\int_V zq\mathrm{d}V}{\int_V q\mathrm{d}V},$$

式中 q 为单位体积重量. 对匀质体，q 为常量，公式化为

$$x_c = \frac{\int_V x\mathrm{d}V}{V},\quad y_c = \frac{\int_V y\mathrm{d}V}{V},\quad z_c = \frac{\int_V z\mathrm{d}V}{V},$$

式中 V 为体积. 这时重心只取决于物体的形状和大小. 此时的重心称为形心. 对匀质曲面，重心 c 的坐标为

$$x_c = \frac{\int_S x\mathrm{d}S}{S},\quad y_c = \frac{\int_S y\mathrm{d}S}{S},\quad z_c = \frac{\int_S z\mathrm{d}S}{S},$$

式中 S 为曲面面积. 对匀质曲线段，重心 c 的坐标为

$$x_c = \frac{\int_l x\mathrm{d}l}{l},\quad y_c = \frac{\int_l y\mathrm{d}l}{l},\quad z_c = \frac{\int_l z\mathrm{d}l}{l},$$

式中 l 为曲线长度. 在实用上，重心位置往往影响到安全和稳定.

摩擦(friction)　一种重要的力学现象. 是互相接触的物体在接触处阻碍相对运动的现象. 该阻碍的力或力偶称为摩擦力或摩阻力偶. 其有几种分类法. 按接触物体的相对运动状态分为静摩擦(由发生相对运动趋势产生)和动摩擦(由相对运动产生). 按接触物体运动形式分为滑动摩擦(由相对滑动(含滑动趋势)引起)和滚动摩擦(由相对滚动(含滚动趋势)引起). 按接触面润滑情况分为干摩擦(接触面纯净)、流体摩擦(接触面有一层液体或气体润滑膜，摩擦发生在流体内部)、边界摩擦(表面有不大于 0.01μm 厚的边界膜所发生的摩擦)和混合膜摩擦(接触面同时存在干摩擦和边界摩擦，或者同时存在流体摩擦和边界摩擦的情况). 在工程上和日常生活中，摩擦问题非常重要，某些情况需要利用摩擦，有些情况又需要避免摩擦，而这些均需依据有关理论处理相应的问题. 摩擦理论涉及多种学科，力学中不探究其物理原因，只讨论其力学现象.

摩擦力(friction force)　见“摩擦”.

静摩擦(static friction)　见“摩擦”.

动摩擦(kinetic friction)　见“摩擦”.

滑动摩擦(sliding friction)　摩擦的一类. 指相互接触的物体，当有相对滑动或相对滑动趋势时，在接触处阻碍相对滑动的现象，称该阻碍力为滑动摩擦力. 接触表面相对静止但有滑动趋势，其摩擦和摩擦力分别称为静滑动摩擦和静滑动摩擦力；接触表面相对滑动，其摩擦和摩擦力分别称为动滑动摩擦和动滑动摩擦力. 滑动摩擦力与相对滑动以及相对滑动的趋势方向相反. 若记动、静二滑动摩擦力分别为 $F_{动}$ 和 $F_{静}$，按通常应用的干摩擦机械啮合论，有

$$F_{动} = \mu' N,\quad 0 \leqslant F_{静} \leqslant F_{\max} = \mu N,$$

式中 N 为接触处的正压力，μ'，μ 分别为动、静摩擦系数. 静滑动摩擦力的最大值 $F_{\max} = \mu N$ 是库伦(Coulomb，C. A. de)经大量实验得出的，称为库伦摩擦定律.

滑动摩擦力(sliding friction force)　见“滑动摩擦”.

库伦摩擦定律(Coulomb law of friction)　见“滑动摩擦”.

滚动摩阻(rolling friction resistance)　亦称滚动摩擦. 摩擦的一类. 指一物体在另一物体表面作相对滚动或有相对滚动趋势时产生的阻碍. 该阻碍作用由力偶矩 M 来度量，其大小有一定范围，即

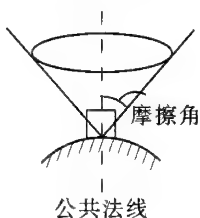
$$0 \leqslant M \leqslant M_{\max} = \delta N,$$

式中 N 为正压力，δ 为滚动摩阻系数，可由实验测定. δ 与材料性质及接触面情况有关，具有长度量纲，单位一般用 mm 或 cm. 最大摩阻力偶矩 $M_{\max}=\delta N$ 称为滚动摩阻定律.

滚动摩擦(rolling friction)　即"滚动摩阻".

滚动摩阻定律(law of rolling friction resistance)　见"滚动摩阻".

摩擦角(angle of friction)　力学的基本概念之一. 指以两物体接触面的公共法线为一边的角，即两物体之间全约束反力的最大值和接触面公共法线的夹角，其正切等于相应的静摩擦系数. 以该公共法线为轴，以 2 倍摩擦角为顶角的圆锥称为摩擦锥. 设这两物体中的一个为约束体，另一个所受全部主动力的合力使其受压且作用线在锥内，则二物体必平衡；若二者作用线在锥外，二物体必失平衡. 二者均不依赖力的大小，称这种平衡现象为自锁. 实践中，有时需利用自锁，如锁紧螺钉，有时要避免自锁，如启闭抽屉，都可用摩擦角的性质来分析.



摩擦锥(cone of friction)　见"摩擦角".

自锁(self lock)　见"摩擦角".

运动学(kinematics)　力学研究的内容之一. 是研究物体机械运动几何性质的科学. 它不考虑导致运动状态发生变化的原因. 在一般力学中，主要讨论质点、刚体及其系统的运动. 称运动的质点为动点，运动学研究其轨迹、运动方程、速度和加速度. 运动学理论是科学技术的重要基础，在机器设计中有直接的应用. 在力学发展的早期，运动学只作为动力学的一部分而加以研究. 19 世纪初，随着机械工程的发展，引出了大量运动学的专门问题，即需要从纯几何方面研究物体的机械运动规律，这才使运动学成为独立的研究对象.

参考系(reference system)　亦称为参照系. 力学的基本概念之一. 是与参考体固连的无限大的空间，可用坐标系表示. 物体的机械运动需以另一物体作标准来确定，此标准物体即参考体. 称适合牛顿运动定律的参考系为惯性系，否则为非惯性系.

参照系(reference system)　即"参考系".

自然法(natural method)　亦称弧坐标法. 研究质点运动的一种方法. 结合运动轨迹建立弧坐标从而研究动点运动的方法. 在轨迹上任取一点 O 为原点，规定某一延伸方向为正，另一延伸方向为负，令动点 P 的弧坐标为 s，s 的绝对值等于 P 到 O 的弧长，s 的正负号由 P 在正的或负的一侧而定. s 是时间 t 的连续函数，称 $s=s(t)$ 为弧坐标形式的运动方程. 这种方法描述动点的速度、加速度，可呈现明显的几何意义.

弧坐标(arc coordinate)　参见"自然法".

自然轴系(trihedral axes of space curve)　一种坐标系. 是依曲线的弯挠自然趋势建立的活动坐标架. 曲线上任一点的切向单位向量为 $\boldsymbol{\tau}$，指向弧坐标的正向，主法向单位向量为 $\boldsymbol{n}$，指向曲率中心，令 $\boldsymbol{b}=\boldsymbol{\tau}\times\boldsymbol{n}$ 为次法向单位向量，这三者构成的右手坐标系即自然轴系. $\boldsymbol{\tau}$ 与 $\boldsymbol{n}$ 决定的平面为密切平面，亦称曲率平面；$\boldsymbol{n}$ 与 $\boldsymbol{b}$ 决定的平面为法平面；$\boldsymbol{b}$ 与 $\boldsymbol{\tau}$ 决定的平面为从切面，亦称直切面. 动点的速度和切向加速度均与 $\boldsymbol{\tau}$ 平行，动点的法向加速度沿 $\boldsymbol{n}$ 的方向，速度和加速度均和 $\boldsymbol{b}$ 垂直.

曲率平面(curvature plane)　见"自然轴系".

直切面(rectifying plane)　见"自然轴系".

向径(radius vector)　表明动点在空间位置的向量. 由参考系中某定点到动点连成的有向线段表示. 它一般是时间的向量函数.

速度(velocity)　运动学的基本概念之一. 是动点的向径对时间的变化率. 它表征该点位置变化的快慢及走向，是向量. 设动点的向径为 $\boldsymbol{r}=\boldsymbol{r}(t)$，则速度

$$\boldsymbol{v}=\lim_{\Delta t\to 0}\frac{\Delta\boldsymbol{r}}{\Delta t}=\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t},$$

称 $\Delta\boldsymbol{r}/\Delta t$ 为动点在时间 t 到 $t+\Delta t$ 内的平均速度，$\mathrm{d}\boldsymbol{r}/\mathrm{d}t$ 为 t 时刻的瞬时速度，简称速度. 速度的绝对值称为速率. 速度在弧坐标系、直角坐标系、柱坐标系和球坐标系中分别表示为：

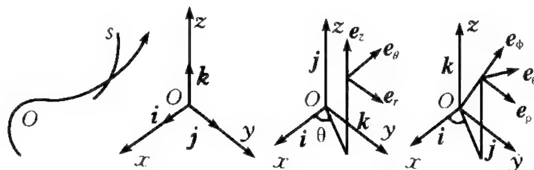
$$\boldsymbol{v}=\frac{\mathrm{d}s}{\mathrm{d}t}\boldsymbol{\tau},$$

$$\boldsymbol{v}=\frac{\mathrm{d}x}{\mathrm{d}t}\boldsymbol{i}+\frac{\mathrm{d}y}{\mathrm{d}t}\boldsymbol{j}+\frac{\mathrm{d}z}{\mathrm{d}t}\boldsymbol{k},$$

$$\boldsymbol{v}=\frac{\mathrm{d}\rho}{\mathrm{d}t}\boldsymbol{e}_\rho+\rho\frac{\mathrm{d}\varphi}{\mathrm{d}t}\boldsymbol{e}_\varphi+\frac{\mathrm{d}z}{\mathrm{d}t}\boldsymbol{k},$$

$$\boldsymbol{v}=\frac{\mathrm{d}r}{\mathrm{d}t}\boldsymbol{e}_r+r\frac{\mathrm{d}\theta}{\mathrm{d}t}\boldsymbol{e}_\theta+r\sin\theta\frac{\mathrm{d}\varphi}{\mathrm{d}t}\boldsymbol{e}_\varphi,$$

式中 $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}$ 分别是 x,y,z 轴上的单位矢量，$\boldsymbol{e}_\rho,\boldsymbol{e}_\varphi,\boldsymbol{e}_r,\boldsymbol{e}_\theta$ 分别是 ρ 方向、φ 增加方向、r 方向、θ 增加方向的单位向量(见图). 在柱坐标系的表达式中，把 z 轴去掉就得速度在极坐标系中的表达式. 速度的单位，用国际单位制为 m/s 或 cm/s，日常生活中常用km/h.



平均速度(average velocity)　见"速度".

瞬时速度(instantaneous velocity)　见"速度".

速率(rate of velocity)　见"速度".

加速度(acceleration) 运动学的基本概念之一. 是动点的速度对时间的变化率,或动点的向径对时间的二阶导数. 加速度

$$\boldsymbol{a}=\frac{\mathrm{d}\,\boldsymbol{v}}{\mathrm{d}t}=\frac{\mathrm{d}^2\,\boldsymbol{r}}{\mathrm{d}t^2},$$

$\boldsymbol{v}$,$\boldsymbol{r}$ 分别为动点的速度和向径. 加速度表征速度随时间变化的情况,可沿动点的轨迹分解为切向和法向两个分量,即 $\boldsymbol{a}=a_\tau\boldsymbol{\tau}+a_n\boldsymbol{n}$,$\boldsymbol{\tau}$ 为切向单位向量,$\boldsymbol{n}$ 为主法向单位向量,

$$a_\tau=\frac{\mathrm{d}\,|\boldsymbol{v}|}{\mathrm{d}t}$$

反映速度大小的变化,$a_n=\boldsymbol{v}^2/\rho$ 反映速度方向的变化,ρ 为轨道的曲率半径. 在直角坐标系、柱坐标系和球坐标系中,加速度的表达式分别为

$$\boldsymbol{a}=\frac{\mathrm{d}^2x}{\mathrm{d}t^2}\boldsymbol{i}+\frac{\mathrm{d}^2y}{\mathrm{d}t^2}\boldsymbol{j}+\frac{\mathrm{d}^2z}{\mathrm{d}t^2}\boldsymbol{k},$$

$$\boldsymbol{a}=\left[\frac{\mathrm{d}^2\rho}{\mathrm{d}t^2}-\rho\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2\right]\boldsymbol{e}_\rho+\frac{1}{\rho}\frac{\mathrm{d}}{\mathrm{d}t}\left(\rho^2\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)\boldsymbol{e}_\varphi+\frac{\mathrm{d}^2z}{\mathrm{d}t^2}\boldsymbol{k},$$

$$\begin{aligned}\boldsymbol{a}=&\left[\frac{\mathrm{d}^2r}{\mathrm{d}t^2}-r\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2-r\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2\sin^2\theta\right]\boldsymbol{e}_r\\&+\left[r\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}+2\frac{\mathrm{d}r}{\mathrm{d}t}\frac{\mathrm{d}\theta}{\mathrm{d}t}-r\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2\sin\theta\cos\theta\right]\boldsymbol{e}_\theta\\&+\left[r\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2}\sin\theta+2\frac{\mathrm{d}r}{\mathrm{d}t}\frac{\mathrm{d}\varphi}{\mathrm{d}t}\sin\theta+2r\frac{\mathrm{d}\varphi}{\mathrm{d}t}\frac{\mathrm{d}\theta}{\mathrm{d}t}\cos\theta\right]\boldsymbol{e}_\varphi.\end{aligned}$$

把柱坐标系中表达式中的 z 去掉,就得到加速度在极坐标系中的表达式. 在惯性参考系中,物体所受外力的主向量除以质量,就等于它质心的加速度. 在地球表面,物体受地球引力产生的加速度减去随地球自转的向心加速度,就等于一般所谓的重力加速度,即自由下落时的加速度(忽略空气阻力),其大小记为 g,其指向偏离地心的程度及其数值,在不同地区略有不同,为便于计算,常近似地取为 $9.8\mathrm{m/s^2}$. 精确地测量重力加速度,在科学技术上有重大意义. 加速度的单位,用国际单位制为 $\mathrm{m/s^2}$,或 $\mathrm{cm/s^2}$.

绝对运动(absolute motion) 运动学的基本概念之一. 指物体相对固定参考系的运动,是为分析复杂运动而建立的概念(参见"点的合成运动").

相对运动(relative motion) 运动学的基本概念之一. 指物体相对运动参考系的运动.

牵连运动(convected motion) 运动学的基本概念之一. 指运动坐标系相对固定坐标系的运动. 是研究物体合成运动时引入的概念(参见"点的合成运动").

绝对速度(absolute velocity) 运动学的基本概念之一. 指动点相对固定坐标系的速度(参见"点的合成运动").

牵连速度(convected velocity) 运动学的基本概念之一. 指动点在运动坐标系上的瞬时重合点相对固定坐标系的速度. 是研究点的合成运动时引入的概念(参见"点的合成运动").

绝对加速度(absolute acceleration) 运动学的基本概念之一. 指动点相对固定坐标系的加速度(参见"点的合成运动").

相对速度(relative velocity) 运动学的基本概念之一. 指动点相对运动坐标系的速度(参见"点的合成运动").

相对加速度(relative acceleration) 运动学的基本概念之一. 指动点相对运动坐标系的加速度(参见"点的合成运动").

牵连加速度(convected acceleration) 运动学的基本概念之一. 指动点在运动坐标系上的瞬时重合点相对固定坐标系的加速度. 研究点的合成运动时引入的概念(参见"点的合成运动").

科里奥利加速度(Coriolis acceleration) 亦称为附加加速度. 简称科氏加速度. 一种加速度. 运动参考系有转动角速度,动点相对运动参考系有相对速度,它们相互影响而产生的加速度. 因法国工程师科里奥利(Coriolis,G. G. de)于 1832 年首先发现而得名. 若 $\boldsymbol{\omega}$ 为运动参考系的角速度,$\boldsymbol{v}_r$ 为动点的相对速度,则科氏加速度为

$$\boldsymbol{a}_k=2\boldsymbol{\omega}\times\boldsymbol{v}_r,$$

该公式亦称为科氏定理. 当 $\boldsymbol{\omega}$,$\boldsymbol{v}_r$ 相互平行或有一个为 0,才使 $\boldsymbol{a}_k$ 为 0. 例如,在北半球,若河水由北向南,而地球角速度表示为向量,由南极指向北极,河水的科氏加速度就由西向东,这样必受到西岸的作用力,西岸就因反作用力而遭冲刷. 据此公式可知,若观察者面向流速方向,北半球南北向的河流总会因此力而使右岸受冲刷.

附加加速度(added acceleration) 即"科里奥利加速度".

科里奥利定理(Coriolis theorem) 见"科里奥利加速度".

点的合成运动(compound motion of particle) 亦称点的复合运动. 质点简单运动的复合,即动点相对固定坐标系的运动、相对运动坐标系的运动以及运动坐标系本身的运动三者存在关系,通过相对运动和运动坐标系的运动来求质点相对固定坐标系的运动称为运动的合成. 主要说明速度的合成关系和加速度的合成关系. 从具有相对运动的二坐标系考察同一点的运动,必不相同. 不妨设其中一个坐标系为定系,另一个为动系,称动点相对定系的运动为绝对运动,相对定系的速度(加速度)为绝对速度 $\boldsymbol{v}_a$(绝对加速度 $\boldsymbol{a}_a$),动点相对动系的运动为相对运动,相对动系的速度(加速度)为相对速度 $\boldsymbol{v}_r$(相对加速度 $\boldsymbol{a}_r$);动系相对定系的运动为牵连运动,动系上动点瞬时重合点相对定系的速度(加速度)为牵连速度 $\boldsymbol{v}_e$(牵连加速度 $\boldsymbol{a}_e$). 关系式

$$\boldsymbol{v}_a = \boldsymbol{v}_e + \boldsymbol{v}_r \text{ 和 } \boldsymbol{a}_a = \boldsymbol{a}_e + \boldsymbol{a}_r + \boldsymbol{a}_k$$

分别表示速度合成定理和加速度合成定理. 式中 $\boldsymbol{a}_k = 2\boldsymbol{\omega} \times \boldsymbol{v}_r$ 是科里奥利加速度, $\boldsymbol{\omega}$ 是动系相对定系转动的角速度,这就可把复杂的运动看成是简单运动的复合,能有效地分析复杂机构的运动.

点的复合运动(compound motion of particle) 即"点的合成运动".

速度合成定理(compound theorem of velocity) 见"点的合成运动".

加速度合成定理(compound theorem of acceleration) 见"点的合成运动".

刚体平行移动(translation motion of rigid body) 刚体运动的形式之一. 指刚体上任一线段保持与其原位平行的运动. 简称刚体平动. 在运动过程中,其上任两点的速度相同,任两点的加速度相同,任两点的轨迹只相当于互相平移一个距离. 这种问题的研究可归结为研究其上任一点的运动,是刚体的简单运动,是研究刚体复杂运动的基础.

刚体绕定轴转动(rotation of rigid body about fixed axis) 刚体运动的形式之一. 指刚体连带其延伸空间上有一直线为固定不动的运动. 称该直线为轴线或转轴. 确定刚体位置的参量是转角,即刚体上包含转轴的平面转过的角,记为 φ,可规定某一转向为正,另一转向为负. 称转角的改变量 $\Delta\varphi$ 为角位移, φ 是时间 t 的单值连续函数,称 $\varphi = \varphi(t)$ 为转动方程,

$$\omega = \frac{\mathrm{d}\varphi}{\mathrm{d}t}$$

为角速度,

$$\varepsilon = \frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2}$$

为角加速度. φ 的单位常用 rad, ω 和 ε 的单位分别为 rad/s 和 rad/s^2. 刚体上转轴以外的点,都作圆心在轴上的圆运动,其速度等于 ωr;切向加速度等于 εr,方向为沿圆的切线;法向加速度等于 $\omega^2 r$,指向转轴, r 为该点到转轴的距离. 其运动微分方程为

$$J\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2} = L,$$

L 是所受主动力对转轴力矩的代数和, J 是刚体对转轴的转动惯量. 运动中,转轴常受到工程上极力避免的附加动压力,只有当转轴为中心惯量主轴时,才能使该附加动压力消失. 对这种运动的研究,是进一步研究刚体复杂运动的基础,其理论也直接用于工程实际.

角位移(angular displacement) 见"刚体绕定轴转动".

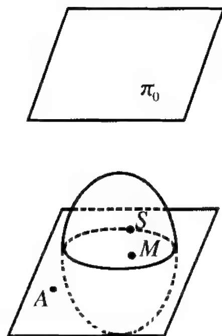
角速度(angular velocity) 运动学的基本概念之一. 是表征物体转动快慢和转向的量. 其大小以转动角度对时间的变化率表示,即角速度

$$\omega = \frac{\mathrm{d}\varphi}{\mathrm{d}t},$$

式中 φ 为转角, t 为时间;其方向用右手螺旋法则确定的方向来表示. 这种有大小和方向的量,也符合平行四边形的合成法则(参见"刚体绕相交轴转动的合成"). 但在镜面反射的坐标变换中,却不能保持其物理特性,所以称它为轴向量(或伪向量、赝向量),常用 $\boldsymbol{\omega}$ 来表示. 对于刚体而言,角速度是滑动轴向量,其所在线,即每一瞬时转动所绕的轴线.

角加速度(angular acceleration) 见"刚体绕定轴转动".

刚体平面运动(plane motion of rigid body) 刚体运动的形式之一. 指刚体上任一点总是与某固定平面保持一定距离的运动. 如图,设固定平面为 π_0,取一平行于 π_0 的平面,在刚体上所截图形 S 的运动规律就能代表整个刚体的情况,故亦称这种运动为平面平行运动. 进一步可看做是平面图形在自身平面的运动. 若在运动图形上取一点 A 称为基点,则图形的运动可看做是随 A 平动和绕 A 转动两部分的合成. 平动情况与 A 的位置有关,转动情况与 A 的位置无关. 运动图形的位置可由 A 的坐标 (x_A, y_A) 和绕 A 的转角 φ 三个参量确定. 运动方程可表为



$$\begin{cases} x_A = x_A(t), \\ y_A = y_A(t), \\ \varphi = \varphi(t). \end{cases}$$

图上任一点 M 的速度、加速度依次为

$$\boldsymbol{v}_M = \boldsymbol{v}_A + \boldsymbol{\omega} \times \overrightarrow{AM},$$

$$\boldsymbol{a}_M = \boldsymbol{a}_A + \boldsymbol{\varepsilon} \times \overrightarrow{AM} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \overrightarrow{AM}),$$

式中 $\boldsymbol{\omega}, \boldsymbol{\varepsilon}$ 分别为用向量表示的图形角速度、角加速度. 称这种运动分析法为基点法. 对于 $\boldsymbol{\omega} \neq 0$ 的每一瞬时,图形连带其延伸平面上必惟一存在速度为 0 的点 P,称为瞬时速度中心(或瞬时转动中心、速度瞬心),则 $\boldsymbol{v}_M = \boldsymbol{\omega} \times \overrightarrow{PM}$. P 在固定平面和运动平面的轨迹分别称为定瞬心轨迹(或空间极迹)和动瞬心轨迹(或本体极迹), P 为二者的切点. 在运动过程中,动瞬心轨迹无滑动地在定瞬心轨迹上滚动. 对于 $\boldsymbol{\omega} \neq 0, \boldsymbol{\varepsilon} \neq 0$ 的每一瞬时,必惟一存在加速度为 0 的点 Q,一般不与 P 重合. 图形上任一点 M 的加速度

$$\boldsymbol{a}_M = \boldsymbol{\varepsilon} \times \overrightarrow{QM} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \overrightarrow{QM}).$$

这种利用速度瞬心和加速度瞬心分析运动的方法称

为瞬心法，工程上常用于机构的运动学分析. 刚体平面运动的动力学微分方程可表示为

$$m\frac{d^2x_c}{dt^2}=\sum X,\ m\frac{d^2y_c}{dt^2}=\sum Y,\ J_c\frac{d^2\varphi}{dt^2}=L_c,$$

式中 m 为刚体的质量，(x_c,y_c) 为质心 c 的坐标，X,Y 为刚体所受外力在 x,y 轴上的投影，J_c 为刚体相对垂直于固定平面且通过 c 的轴的转动惯量，L_c 为所受外力对 c 之矩. 这就需刚体所受外力能简化位于通过质心 c 而平行于 π_0 的平面之内，各点初速度也都平行此平面. 工程上很多运动部件都满足此条件，故呈现平面运动.

基点法(method of base point)　见"刚体平面运动".

平面平行运动(plane parallel motion)　见"刚体平面运动".

瞬心法(method of instantaneous center)　见"刚体平面运动".

瞬时速度中心(instantaneous center of velocity)　见"刚体平面运动".

瞬时转动中心(instantaneous center of rotation)　见"刚体平面运动".

瞬时加速度中心(instantaneous center of acceleration)　见"刚体平面运动".

瞬心轨迹(centerode)　见"刚体平面运动".

空间极迹(herpolhode)　见"刚体平面运动".

本体极迹(polhode)　见"刚体平面运动".

自由刚体的一般运动(general motion of free rigid body)　刚体运动的形式之一. 指刚体不受限制的运动. 在刚体上任取一点 A 为基点，过 A 取一个平动坐标系，刚体的运动可分解为随 A 点的平动(由该坐标系代表)和绕 A 点的转动(相对于平动坐标系的运动)两部分. 刚体的位置可由点 A 在固定坐标系 $O\xi\eta\zeta$ 中的坐标(ξ_A,η_A,ζ_A)和相对平动坐标系的欧拉角 ψ,φ,θ 六个参数确定，运动方程可表示为

$$\xi_A=\xi_A(t),\ \eta_A=\eta_A(t),\ \zeta_A=\zeta_A(t),$$
$$\psi=\psi(t),\ \varphi=\varphi(t),\ \theta=\theta(t).$$

刚体上任一点 M 的速度 $\boldsymbol{v}_M$ 和加速度 $\boldsymbol{a}_M$ 都可按合成法表示：

$$\boldsymbol{v}_M=\boldsymbol{v}_A+\boldsymbol{\omega}\times\overrightarrow{AM},$$
$$\boldsymbol{a}_M=\boldsymbol{a}_A+\boldsymbol{\varepsilon}\times\overrightarrow{AM}+\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\overrightarrow{AM}),$$

式中 $\boldsymbol{v}_A,\boldsymbol{a}_A$ 分别为 A 的速度和加速度，$\boldsymbol{\omega},\boldsymbol{\varepsilon}$ 分别为用向量表示的刚体角速度、角加速度，$\boldsymbol{v}_A,\boldsymbol{a}_A$ 与基点位置有关，$\boldsymbol{\omega},\boldsymbol{\varepsilon}$ 与基点位置无关. 这种一般运动是一个随时间而变的螺旋运动：刚体一面沿一轴平动，一面绕该轴转动，$\boldsymbol{\omega}$ 即是沿此轴的. 若选基点在此轴线上，则移动速度就等于基点速度. 令 $O\xi\eta\zeta$ 固定于惯性空间，过刚体质心 c 取重合于中心惯性主轴的坐标 $c(x,y,z)$. 则动力学方程可表示为

$$m\frac{d^2\xi_c}{dt^2}=\sum F_\xi,$$
$$m\frac{d^2\eta_c}{dt^2}=\sum F_\eta,$$
$$m\frac{d^2\zeta_c}{dt^2}=\sum F_\zeta,$$
$$J_1\frac{d\omega_1}{dt}+(J_3-J_2)\omega_2\omega_3=L_1,$$
$$J_2\frac{d\omega_2}{dt}+(J_1-J_3)\omega_3\omega_1=L_2,$$
$$J_3\frac{d\omega_3}{dt}+(J_2-J_1)\omega_1\omega_2=L_3,$$

式中：m 是刚体质量；(ξ_c,η_c,ζ_c)是质心坐标；F_ξ,F_η,F_ζ 分别是刚体上外力在 ξ,η,ζ 上的投影；J_1,J_2,J_3 分别是刚体的中心主转动惯量；$\omega_1,\omega_2,\omega_3,L_1,L_2,L_3$ 分别是刚体角速度和外力对 c 点的主矩在 x,y,z 的投影. 一般情况下求解，还需和欧拉运动学方程(参见"欧拉运动学方程")联立起来.

刚体螺旋运动(helical motion of rigid body)　刚体运动的形式之一. 指刚体绕某一轴线转动，同时又沿该轴平动. 若用矢量表示转动的角速度，当它与平动速度同向，称为右螺旋，异向称为左螺旋. 一般情况下，自由刚体在每一瞬时的运动，都可看做是螺旋运动. 这是因为刚体的一般运动总可分解为随基点平动和绕基点转动两部分，在刚体连带其延伸空间上适当选取基点，就可以使平动速度和转轴平行.

刚体绕平行轴转动的合成(composition of rotation of rigid body about parallel axis)　一种刚体运动. 指刚体由两个或两个以上转动运动的简化. 刚体以大小为 ω_1 的角速度绕轴 1 转动，轴 1 又以大小为 ω_2 的角速度绕平行于轴 1 的轴 2 转动，若二者转向相同，则刚体必以大小为 $\omega_1+\omega_2$ 的角速度转动，其瞬时转轴和原二轴平行，位置是以二角速度大小的反比内分二轴的距离；若二角速度转向不同且大小不等，则刚体角速度的大小等于二者大小之差，方向与较大的一个相同，瞬时轴在较大一个角速度所绕轴的外侧，以二角速度大小的反比外分二轴的距离；若二角速度大小相等，转向不同，则形成转动偶，刚体必在垂直于轴线的平面内平动，速度等于一个角速度的大小与其偶矩(二轴的距离)的乘积.

转动偶(rotation couple)　见"刚体绕平行轴转动的合成".

刚体绕相交轴转动的合成(composition of rotation of rigid body about intersection axis)　一种刚体运动. 指刚体由两个或两个以上转动运动的简化. 刚体以角速度 $\boldsymbol{\omega}_1$ 绕 $\boldsymbol{\omega}_1$ 所在轴线 1 转动，轴 1 又以角速度 $\boldsymbol{\omega}_2$ 绕 $\boldsymbol{\omega}_2$ 所在轴线 2 转动，两轴交于 O 点，

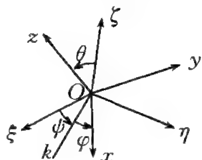
ω_1 和 ω_2 可按向量平行四边形法则合成为 $\omega=\omega_1+\omega_2$，刚体必以角速度 ω 绕 ω 所在轴线转动. 两个以上相交轴转动的合成可依此类推.

刚体绕定点运动(motion of rigid body about fixed point)　刚体运动的形式之一. 指刚体连带其延伸空间上有一固定点的运动. 每一瞬时，刚体总是绕着过该定点的一个变动轴线转动. 称这轴为瞬时转动轴，简称瞬时轴. 该轴在固定空间和运动空间的轨迹，都是一个以固定点为顶点的锥面，分别称为定瞬轴锥面(也称为空间极面)和动瞬轴锥面(也称为本体极面). 瞬时轴即这二锥面的公共母线. 动瞬轴锥面无滑动地在定瞬轴锥面上滚动. 若用 $\boldsymbol{\omega}$ 和 $\boldsymbol{\varepsilon}$ 分别表示刚体的角速度和角加速度，记固定点为 O，刚体上任一点 M 的速度、加速度分别为

$$\boldsymbol{v}=\boldsymbol{\omega}\times\overrightarrow{OM},$$

$$\boldsymbol{a}=\boldsymbol{\varepsilon}\times\overrightarrow{OM}+\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\overrightarrow{OM}).$$

令 ψ,θ,φ 分别表示刚体的进动角、章动角、自转角，合称为欧拉角，刚体在空间的位置可由这三者惟一确定. 如图，$O\xi\eta\zeta$ 为固定坐标系，$Oxyz$ 为随刚体而动的运动坐标系. Ok 为 xy 与 $\xi\eta$ 二平面的交线，ψ 为 ξ 轴到 Ok 的夹角，θ 为 ζ 轴到 z 轴的夹角，φ 为 Ok 到 x 轴的夹角. 它们依赖时间 t 的函数关系



$$\psi=\psi(t),\quad \theta=\theta(t),\quad \varphi=\varphi(t)$$

称为刚体的运动方程. 称

$$\omega_1=\frac{d\psi}{dt}\sin\theta\sin\varphi+\frac{d\theta}{dt}\cos\varphi,$$

$$\omega_2=\frac{d\psi}{dt}\sin\theta\cos\varphi-\frac{d\theta}{dt}\sin\varphi,$$

$$\omega_3=\frac{d\psi}{dt}\cos\theta+\frac{d\varphi}{dt}$$

为欧拉运动学方程；

$$\frac{dG_1}{dt}+\omega_2G_3-\omega_3G_2=L_1,$$

$$\frac{dG_2}{dt}+\omega_3G_1-\omega_1G_3=L_2,$$

$$\frac{dG_3}{dt}+\omega_1G_2-\omega_2G_1=L_3$$

为推广的欧拉运动学方程. 式中 $G_1,G_2,G_3,\omega_1,\omega_2,\omega_3,L_1,L_2,L_3$ 依次为刚体对 O 点的动量矩 $\boldsymbol{G}$、角速度 ω、所受外力对 O 的主矩 $\boldsymbol{L}$ 分别在 x,y,z 轴的投影. 若取 x,y,z 为刚体对 O 的惯量主轴，J_1,J_2,J_3 为相应的三个主转动惯量，方程

$$J_1\frac{d\omega_1}{dt}+\omega_2\omega_3(J_3-J_2)=L_1,$$

$$J_2\frac{d\omega_2}{dt}+\omega_1\omega_3(J_1-J_3)=L_2,$$

$$J_3\frac{d\omega_3}{dt}+\omega_1\omega_2(J_2-J_1)=L_3$$

为瑞士数学家欧拉(Euler,L.)于 1765 年推出，称为欧拉动力学方程. 从理论上讲，若知所受外力和初始条件，就可由欧拉运动学方程和动力学方程联立而确定刚体的运动规律. 18－19 世纪，它曾吸引许多数学家、力学家作了大量工作，以寻求其解析解，但至今只在刚体仅受重力作用下的三种特殊情形，即欧拉情形、拉格朗日情形、柯瓦列夫斯卡娅情形下，才能针对任意的初始条件获得解析解. 20 世纪初，胡松(Husson)、波加提(Burgatti)等人的研究表明，对于任意的初始条件(不是某些特殊的初始条件)，上述三种情形之外不可能再有解析解. 因为刚体的一般运动可分解为随基点平动和绕基点转动两部分，故刚体绕定点运动的理论是研究刚体一般运动的基础. 在实用上，一切天体和航天器都需要确定其轨道和空间方位的变化规律，该理论可用来解决其方位问题；应用于技术领域的陀螺仪近似理论，则属于该理论的简化.

空间极面(herpolhode)　见“刚体绕定点运动”.

定瞬轴锥面(fixed cones of instantaneous axis)　见“刚体绕定点运动”.

本体极面(polhode)　见“刚体绕定点运动”.

动瞬轴锥面(motive cone of instantaneous axis)　见“刚体绕定点运动”.

欧拉动力学方程(Euler dynamical equation)　见“刚体绕定点运动”.

欧拉定理(Euler theorem)　亦称达朗贝尔定理. 研究刚体运动的重要定理. 是刚体绕定点运动的位移定理：具有固定点的刚体的任何有限位移，总可绕过该点的某一轴转动一次而实现. 本定理由法国数学家、力学家达朗贝尔(d'Alembert,J. le R.)于 1749 年提出，瑞士数学家欧拉(Euler,L.)于 1750 年提出，通常多以欧拉命名. 例如从此定理出发，可推出刚体绕定点运动的任一瞬时都存在着瞬时转动轴和瞬时角速度，且这种角速度可按向量的加法合成.

达朗贝尔定理(d'Alembert theorem)　即“欧拉定理”.

里瓦斯公式(Rivals formula)　刚体运动的速度公式. 刚体上任一点绕定点 O 运动时，其上任一点 P 的速度 $\boldsymbol{v}$、加速度 $\boldsymbol{a}$ 的表达式为

$$\boldsymbol{v}=\boldsymbol{\omega}\times\overrightarrow{OP},\quad \boldsymbol{a}=\boldsymbol{\varepsilon}\times\overrightarrow{OP}+\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\overrightarrow{OP}),$$

式中 $\boldsymbol{\omega},\boldsymbol{\varepsilon}$ 分别为刚体的角速度、角加速度，一般情况下二者不共线. 称 $\boldsymbol{\varepsilon}\times\overrightarrow{OP}$ 为转动加速度(旋转加速度)，它垂直于 $\boldsymbol{\varepsilon}$ 和 $\overrightarrow{OP}$，不与 $\boldsymbol{v}$ 共线；称 $\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\overrightarrow{OP})$ 为向轴加速度，垂直于 $\boldsymbol{\omega}$ 和 $\boldsymbol{v}$，但不沿 P 点轨迹

的主法线.刚体绕定轴转动的速度、加速度公式属于 $\boldsymbol{\omega}$ 与 $\boldsymbol{\varepsilon}$ 共线的特殊情况.

欧拉运动学方程(Euler kinematical equation)描述刚体运动的微分方程.在刚体绕定点运动中,反映角速度和欧拉角关系的方程为

$$\omega_1=\frac{\mathrm{d}\psi}{\mathrm{d}t}\sin\theta\sin\varphi+\frac{\mathrm{d}\theta}{\mathrm{d}t}\cos\varphi,$$

$$\omega_2=\frac{\mathrm{d}\psi}{\mathrm{d}t}\sin\theta\cos\varphi-\frac{\mathrm{d}\theta}{\mathrm{d}t}\sin\varphi,$$

$$\omega_3=\frac{\mathrm{d}\psi}{\mathrm{d}t}\cos\theta+\frac{\mathrm{d}\varphi}{\mathrm{d}t},$$

式中 $\omega_1,\omega_2,\omega_3$ 为角速度在运动坐标系三个轴上的投影.ψ,θ,φ 依次为运动坐标系对固定坐标系的进动角、章动角、自转角.该方程在刚体绕定点运动的研究中有重要地位.

速度投影定理(theorem of projection of velocity) 在运动学中,反映刚体特性的一个重要定理.它反映刚体上各点间的速度关系.一切运动中,刚体内任两点的速度在其连线上的投影相等.

四元数(quaternions) 数系的一种推广.一个实数单位 1 和三个虚数单位 i_1,i_2,i_3 作为基,四个实数作为分量的一种数学模型.可用字母代表.例如,

$$M=\lambda_0+\lambda_1i_1+\lambda_2i_2+\lambda_3i_3,$$

式中诸 λ 为实数.有加法、乘法等自身的运算法则.若把三个虚数单位看做是单位向量,记为 $\boldsymbol{i}_1,\boldsymbol{i}_2,\boldsymbol{i}_3$,则可表示为一个标量加一个向量,即

$$M=\lambda_0+\boldsymbol{\lambda}(\boldsymbol{\lambda}=\lambda_1\boldsymbol{i}_1+\lambda_2\boldsymbol{i}_2+\lambda_3\boldsymbol{i}_3).$$

定义 M 与 $N=\mu_0+\boldsymbol{\mu}$ 作乘法为

$$M\circ N=(\lambda_0\mu_0-\boldsymbol{\lambda}\cdot\boldsymbol{\mu})+(\lambda_0\boldsymbol{\mu}+\mu_0\boldsymbol{\lambda}+\boldsymbol{\lambda}\times\boldsymbol{\mu}),$$

而四元数之间的加法以及四元数与实数的乘法均可借用标量和向量的通常运算解决.代入四元数的乘法关系式可以得到

$$i_1\circ i_1=-\boldsymbol{i}_1\cdot\boldsymbol{i}_1=-1,$$

表明 $\boldsymbol{i}_1$ 是虚数单位并且

$$\boldsymbol{i}_2\circ\boldsymbol{i}_3=-\boldsymbol{i}_3\circ\boldsymbol{i}_2=\boldsymbol{i}_1,$$

其余类推.一般地,$M\circ N\neq N\circ M$,又 $M=\lambda_0+\boldsymbol{\lambda}$ 的共轭为 $\widetilde{M}=\lambda_0-\boldsymbol{\lambda}$.根据标量和向量的运算可知

$$\widetilde{(M+N)}=\widetilde{M}+\widetilde{N},$$

$$\widetilde{(M\circ N)}=\widetilde{N}\circ\widetilde{M},$$

$$M\circ\widetilde{M}=\widetilde{M}\circ M=|M|^2.$$

$|M|$ 是 M 的模,且

$$|M|=\sqrt{\lambda_0^2+\lambda_1^2+\lambda_2^2+\lambda_3^2},$$

而 M 的逆

$$M^{-1}=\widetilde{M}/|M|^2.$$

英国数学家、物理学家哈密顿(Hamilton,W. R.)曾考虑以类似平面上复数处理几何问题的方法来研究空间几何问题,四元数理论就是他于 1843 年以这个思路建立的.在他之后也有人对四元数继续进行研究,但长期未用于实际问题.20 世纪 70 年代,苏联学者运用四元数处理刚体定位问题,从而在刚体运动中得到了应用.主要根据是四元数 M 对另一个四元数 L 的旋转变换 $M\circ L\circ M^{-1}=L'$ 有:

1. $|L'|=|L|$.

2. L' 和 L 的标量部分相等.

3. L' 的向量部分是由 L 的向量部分沿着以 $\boldsymbol{\lambda}$ 为轴的锥面旋转角度 $\varphi=2\arccos(\lambda_0/|M|)$ 而得.

若 $Q=q_0+q_1\boldsymbol{i}_1+q_2\boldsymbol{i}_2+q_3\boldsymbol{i}_3$ 为单位四元数,即 $|Q|=1$,而 $\boldsymbol{r}=r_1\boldsymbol{i}_1+r_2\boldsymbol{i}_2+r_3\boldsymbol{i}_3$ 是标量部分为 0 的四元数,则 Q 把 $\boldsymbol{r}$ 旋转为 $\boldsymbol{r}'$,即

$$Q\circ\boldsymbol{r}\circ Q^{-1}=\boldsymbol{r}'=r_1'\boldsymbol{i}_1+r_2'\boldsymbol{i}_2+r_3'\boldsymbol{i}_3,$$

坐标变换关系为

$$[r_1'\ r_2'\ r_3']^{\mathrm{T}}=A[r_1\ r_2\ r_3]^{\mathrm{T}},$$

右上角字母 T 表示转置,A 为变换矩阵,其元素分别为

$$A_{11}=2(q_0^2+q_1^2)-1,\qquad A_{12}=2(q_1q_2-q_0q_3),$$
$$A_{13}=2(q_3q_1+q_0q_2),\qquad A_{21}=2(q_1q_2+q_0q_3),$$
$$A_{22}=2(q_0^2+q_2^2)-1,\qquad A_{23}=2(q_2q_3-q_0q_1),$$
$$A_{31}=2(q_3q_1-q_0q_2),\qquad A_{32}=2(q_2q_3+q_0q_1),$$
$$A_{33}=2(q_0^2+q_3^3)-1.$$

这就是刚体绕定点运动时,设转轴上的向量为

$$\boldsymbol{q}=q_1\boldsymbol{i}_1+q_2\boldsymbol{i}_2+q_3\boldsymbol{i}_3,$$

按右手螺旋法则发生转角为 $\varphi=2\arccos q_0$,使其上的向径 $\boldsymbol{r}$ 转为 $\boldsymbol{r}'$ 的坐标变换公式.反过来,设刚体绕定点转角为 φ,转轴上单位向量为

$$\boldsymbol{\tau}=\tau_1\boldsymbol{i}_1+\tau_2\boldsymbol{i}_2+\tau_3\boldsymbol{i}_3,$$

则令

$$q_0=\cos\frac{\varphi}{2},\quad q_1=\tau_1\sin\frac{\varphi}{2},$$

$$q_2=\tau_2\sin\frac{\varphi}{2},\quad q_3=\tau_3\sin\frac{\varphi}{2},$$

也可得变换矩阵 A.瑞士数学家欧拉(Euler, L.)就是从这个几何关系得到 q_0,q_1,q_2,q_3 的,故也称这四个 q 为欧拉参数.这种参数与欧拉角、框架角有确定的对应关系,只要比较其变换的矩阵就可得到.而采用这四个数可以避免三角函数运算,也不会出现奇点,且有一定的对称性,便于记忆.缺点是四个参数不独立,不能作为广义坐标(需有一个约束条件).四元数理论使欧拉参数更显出直观性,并使它立足于完美的数学理论体系之上.

欧拉参数(Euler parameter) 描述刚体绕定点

转动状态的四个参数.是刚体绕定点运动时,确定其转动量的四个参数.设刚体绕通过定点的向量 $\boldsymbol{\lambda}=\lambda_1\boldsymbol{i}_1+\lambda_2\boldsymbol{i}_2+\lambda_3\boldsymbol{i}_3$ ($\boldsymbol{i}_1,\boldsymbol{i}_2,\boldsymbol{i}_3$ 是过定点的互相正交的单位向量),依右手螺旋法则的转角为 φ,则令

$$q_0=\cos\frac{\varphi}{2},\qquad q_1=\frac{\lambda_1}{|\boldsymbol{\lambda}|}\sin\frac{\varphi}{2},$$
$$q_2=\frac{\lambda_2}{|\boldsymbol{\lambda}|}\sin\frac{\varphi}{2},\quad q_3=\frac{\lambda_3}{|\boldsymbol{\lambda}|}\sin\frac{\varphi}{2},$$
$$(|\boldsymbol{\lambda}|=\sqrt{\lambda_1^2+\lambda_2^2+\lambda_3^2})$$

这些就是欧拉参数,称 $q_1\boldsymbol{i}_1+q_2\boldsymbol{i}_2+q_3\boldsymbol{i}_3$ 为欧拉向量.可用欧拉参数表示出刚体上任一点转动前后的坐标变换关系,即

$$x'=[2(q_0^2+q_1^2)-1]x+2(q_1q_2-q_0q_3)y+2(q_3q_1+q_0q_2)z,$$
$$y'=2(q_1q_2+q_0q_3)x+[2(q_0^2+q_2^2)-1]y+2(q_2q_3-q_0q_1)z,$$
$$z'=2(q_3q_1-q_0q_2)x+2(q_2q_3+q_0q_1)y+[2(q_0^2+q_3^2)-1]z,$$

(x,y,z), (x',y',z') 分别是同一点在转动前、转动后在固定坐标系下的坐标.欧拉参数与欧拉角、布里恩角都有确定的关系,只需比较各自的变换关系即可得到.也可用欧拉参数表示出刚体的角速度和角加速度,例如,

$$\omega_1=2(\dot{q}_1q_0-\dot{q}_0q_1+q_3\dot{q}_2-q_2\dot{q}_3),$$
$$\omega_2=2(\dot{q}_2q_0-\dot{q}_0q_2+q_1\dot{q}_3-q_3\dot{q}_1),$$
$$\omega_3=2(\dot{q}_3q_0-\dot{q}_0q_3+q_2\dot{q}_1-q_1\dot{q}_2),$$
$$\varepsilon_1=2(-q_1\ddot{q}_0+q_0\ddot{q}_1+q_3\ddot{q}_2-q_2\ddot{q}_3),$$
$$\varepsilon_2=2(-q_2\ddot{q}_0-q_3\ddot{q}_1+q_0\ddot{q}_2+q_1\ddot{q}_3),$$
$$\varepsilon_3=2(-q_3\ddot{q}_0+q_2\ddot{q}_1-q_1\ddot{q}_2+q_0\ddot{q}_3),$$

分别是角速度 $\boldsymbol{\omega}$、角加速度 $\boldsymbol{\varepsilon}$ 在随刚体运动的坐标轴上的投影.各个 q 上的点号表示对时间求导数.用欧拉参数表示的运动学方程

$$\dot{q}_0=\frac{1}{2}(-\omega_1q_1-\omega_2q_2-\omega_3q_3),$$
$$\dot{q}_1=\frac{1}{2}(\omega_1q_0-\omega_2q_3+\omega_3q_2),$$
$$\dot{q}_2=\frac{1}{2}(\omega_1q_3+\omega_2q_0-\omega_3q_1),$$
$$\dot{q}_3=\frac{1}{2}(-\omega_1q_2+\omega_2q_1+\omega_3q_0)$$

不会产生奇点,当 $\omega_1,\omega_2,\omega_3$ 为已知函数时,便于进行数值积分.

欧拉向量(Euler vector)　见"欧拉参数".

罗德里格参数(Rodrigues parameter)　描述刚体绕定点转动的三个参数.设刚体绕过定点的向量 $\boldsymbol{\lambda}=\lambda_1\boldsymbol{i}_1+\lambda_2\boldsymbol{i}_2+\lambda_3\boldsymbol{i}_3$ ($\boldsymbol{i}_1,\boldsymbol{i}_2,\boldsymbol{i}_3$ 是互相正交的过定点的单位向量),按右手螺旋法则转角为 φ,则

$$\rho_1=\frac{\lambda_1}{|\boldsymbol{\lambda}|}\tan\frac{\varphi}{2},\ \rho_2=\frac{\lambda_2}{|\boldsymbol{\lambda}|}\tan\frac{\varphi}{2},\ \rho_3=\frac{\lambda_3}{|\boldsymbol{\lambda}|}\tan\frac{\varphi}{2}$$

$(|\boldsymbol{\lambda}|=\sqrt{\lambda_1^2+\lambda_2^2+\lambda_3^2})$ 为罗德里格参数.称 $\boldsymbol{\rho}=\rho_1\boldsymbol{i}_1+\rho_2\boldsymbol{i}_2+\rho_3\boldsymbol{i}_3$ 为罗德里格向量.与欧拉参数 q_0,q_1,q_2,q_3 有如下关系:

$$q_0=(1+\rho_1^2+\rho_2^2+\rho_3^2)^{-1/2},\quad q_1=\rho_1q_0,$$
$$q_2=\rho_2q_0,\quad q_3=\rho_3q_0.$$

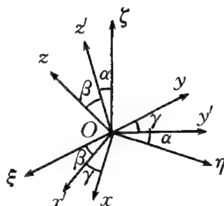
刚体上旋转轴外的任一点在固定坐标系中的坐标,旋转前设为 (ξ,η,ζ),旋转后为 (x,y,z),则变换关系为

$$x=[(1+\rho_1^2-\rho_2^2-\rho_3^2)\xi+2(\rho_1\rho_2-\rho_3)\eta+2(\rho_1\rho_3+\rho_2)\zeta](1+\rho_1^2+\rho_2^2+\rho_3^2)^{-1},$$
$$y=[2(\rho_1\rho_2+\rho_3)\xi+(1+\rho_2^2-\rho_3^2-\rho_1^2)\eta+2(\rho_2\rho_3-\rho_1)\zeta](1+\rho_1^2+\rho_2^2+\rho_3^2)^{-1},$$
$$z=[2(\rho_3\rho_1-\rho_2)\xi+2(\rho_3\rho_2+1)\eta+(1+\rho_3^2-\rho_1^2-\rho_2^2)\zeta](1+\rho_1^2+\rho_2^2+\rho_3^2)^{-1}.$$

这种参数的优点是只有三个参数,但其缺点是当 $\varphi=\pi$ 时,参数无意义.

罗德里格向量(Rodrigues vector)　见"罗德里格参数".

布里恩角(Bryant angle)　亦称框架角,或台特角,或卡尔丹角,或克雷洛夫角.广义欧拉角(参见"欧拉角")是确定刚体方位的一种广义坐标系中的三个角.如图,设 $O\xi\eta\zeta$ 为参考系中的固定坐标系,设想原先固连于刚体的坐标系 $O\xi'\eta'\zeta'$ 与固定坐标系重合,先令刚体绕 ξ 顺时针转 α 角,使 η' 到达 y',ζ' 到达 z',再令刚体绕 y' 顺时针转 β 角,使 ξ' 到达 x',z' 到达 z,最后令刚体绕 z 顺时针转 γ 角,使 x' 到达 x,y' 到达 y. $Oxyz$ 的位置就代表刚体绕 O 顺时针任意转动后到达的位置,α,β,γ 为布里恩角,可作为广义坐标来确定刚体方位.它们绕轴的转向和轴向均符合右手螺旋法则.连体坐标和固定坐标的关系为



$$(\boldsymbol{e}_1\ \boldsymbol{e}_2\ \boldsymbol{e}_3)A=(\boldsymbol{\varepsilon}_1\ \boldsymbol{\varepsilon}_2\ \boldsymbol{\varepsilon}_3),$$

$\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3,\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3$ 依次为 x,y,z,ξ,η,ζ 轴上的单位向量,正交矩阵 A 的元素分别为

$$A_{11}=\cos\beta\cos\gamma,$$
$$A_{12}=\cos\alpha\sin\gamma+\sin\alpha\sin\beta\cos\gamma,$$
$$A_{13}=\sin\alpha\sin\gamma-\cos\alpha\sin\beta\cos\gamma,$$
$$A_{21}=-\cos\beta\sin\gamma,$$
$$A_{22}=\cos\alpha\cos\gamma-\sin\alpha\sin\beta\sin\gamma,$$
$$A_{23}=\sin\alpha\cos\gamma+\cos\alpha\sin\beta\sin\gamma,$$
$$A_{31}=\sin\beta,$$
$$A_{32}=-\sin\alpha\cos\beta,$$
$$A_{33}=\cos\alpha\cos\beta.$$

在奇点 $\beta=\pi/2+n\pi$ $(n=0,1,2,\cdots)$ 附近会发生数

值计算的困难.但当角度接近于零,则很便于研究问题.当三个角都足够小时,矩阵 A 的元素可近似为

$$A_{11}=A_{22}=A_{33}=1,\quad A_{12}=-A_{21}=\gamma,$$
$$A_{31}=-A_{13}=\beta,\qquad A_{23}=-A_{32}=\alpha.$$

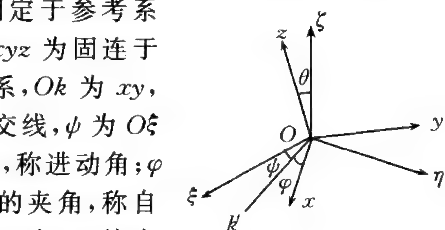
台特角(Tait angle)　即“布里恩角”.

克雷洛夫角(Krulov angle)　即“布里恩角”.

卡尔丹角(Cardan angle)　即“布里恩角”.

框架角(frame angle)　即“布里恩角”.

欧拉角(Euler angle)　经典刚体动力学中习惯使用的广义坐标系中的角.是确定刚体方位的进动角、自转角、章动角的统称.特别适用于分析章动角接近不变,进动角和自转角接近匀速增长的情形.如图,$O\xi\eta\zeta$ 为固定于参考系的坐标系,$Oxyz$ 为固连于刚体的坐标系,Ok 为 xy,$\xi\eta$ 两平面的交线,ψ 为 $O\xi$ 到 Ok 的夹角,称进动角;φ 为 Ok 到 Ox 的夹角,称自转角;θ 为 $O\zeta$ 到 Oz 的夹角,称章动角.设想原先 x,y,z 轴依次重合于 ξ,η,ζ 轴,令刚体绕 $O\zeta$ 顺时针转 ψ 角,则 Ox 到达 Ok 位置,再令刚体绕 Ok 顺时针转 θ 角,则 Oz 到达图示位置,最后令刚体绕 Oz 顺时针转 φ 角,则 Ox,Oy 也达图示位置.坐标变换关系为



$$(\boldsymbol{\varepsilon}_1\ \boldsymbol{\varepsilon}_2\ \boldsymbol{\varepsilon}_3)=(\boldsymbol{e}_1\ \boldsymbol{e}_2\ \boldsymbol{e}_3)A,$$

$\boldsymbol{\varepsilon}_1$,$\boldsymbol{e}_1$ 分别为 ξ 轴、x 轴上的单位向量,其余类推.矩阵 A 的元素为

$$A_{11}=\cos\psi\cos\varphi-\sin\psi\sin\varphi\cos\theta,$$
$$A_{12}=\sin\psi\sin\varphi+\cos\psi\cos\theta\sin\varphi,$$
$$A_{13}=\sin\theta\sin\varphi,$$
$$A_{21}=-\cos\psi\sin\varphi-\sin\psi\cos\theta\cos\varphi,$$
$$A_{22}=-\sin\psi\sin\varphi+\cos\psi\cos\theta\cos\varphi,$$
$$A_{23}=\sin\theta\cos\varphi,\qquad A_{31}=\sin\psi\sin\theta,$$
$$A_{32}=-\cos\psi\sin\theta,\quad A_{33}=\cos\theta.$$

用欧拉角作广义坐标系中的角,当 θ 接近于临界值 $n\pi$ ($n=0,1,\cdots$)时,数值计算将出现困难.

进动角(precession angle)　见“欧拉角”.

自转角(self rotation angle)　见“欧拉角”.

章动角(nutation angle)　见“欧拉角”.

动力学(dynamics)　力学的一个分支.是研究物体机械运动和受力关系的科学.即已知外力和初始条件求运动规律,或由运动确定所受之力.其理论依据主要是牛顿定律.作为理论力学的一个部分,将具体讨论质点、质点系和刚体的动力学问题,为此需进行受力分析和运动学分析,静力学和运动学可为其基础.工程技术领域中存在着丰富的动力学问题,使力学在应用中形成了众多的动力学学科,如断裂动力学、多刚体动力学、大气飞行动力学、结构动力学、塑性动力学、心脏动力学、厚板动力学等.

牛顿定律(Newton laws)　力学的基本定律之一.即牛顿三定律.一般表述为:

1. 惯性定律.质点不受力将永远保持静止或匀速直线运动.

2. 运动定律.质点受力将产生与力同向的加速度,其大小与力的大小成正比,而与质点的质量成反比.

3. 作用与反作用定律.两物体间的作用力与反作用力大小相等、方向相反,沿同一作用线分别作用于该两物体.

牛顿定律由英国数学家、物理学家牛顿(Newton,I.)于 1687 年在《自然哲学的数学原理》中提出,适用于惯性参考系,且速度远小于光速的宏观物体.在日常生活和工程技术中,以地球为参考体应用该定律也能获得足够精确的结果.在与工程实践的结合中,以牛顿定律为理论基础的力学各学科得到飞速的发展.按照相对论力学的动力学方程为

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0\boldsymbol{v}}{\sqrt{1-\frac{|\boldsymbol{v}|^2}{c^2}}}\right)=\boldsymbol{F},$$

式中 m_0 为质点的静止质量,$\boldsymbol{v}$ 为质点速度,$\boldsymbol{F}$ 为质点所受之力,c 为光速.当速度 $\boldsymbol{v}$ 的值远小于光速时,才可视质量等于静止质量而为常量,可得

$$m_0\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=\boldsymbol{F},$$

即为上述第二定律的表达式.

惯性定律(law of inertia)　即牛顿第一定律.见“牛顿定律”.

惯性(inertia)　物体的固有力学性质.指物体固有的保持其运动状态不变的特性.称匀速直线平动为惯性运动.物体惯性的大小,体现在受到外力时速度发生改变(含大小和方向的改变)的难易程度上.受同样的外力,速度改变较难的惯性就大.此概念首先由意大利科学家伽利略(Galilei,G.)提出,英国数学家、物理学家牛顿(Newton,I.)加以完善.根据牛顿第二定律,物体的加速度与所受的力成正比,与质量成反比,质量就是惯性的度量,称这种质量为惯性质量.在国际单位制中,质量的单位为千克.

惯性运动(inertial motion)　见“惯性”.

质量(mass)　反映物质普遍属性的物理量.在物体受力产生加速度的现象中,它度量物体的惯性,称为惯性质量;在引力现象中,它度量物体之间引力的大小,称为引力质量.在英国数学家、物理学家牛顿(Newton,I.)之后,许多科学家所做的精密实验中,特别是 1916 年,美籍德国科学家爱因斯坦(Einstein,A.)的广义相对论,证明了二者完全相同,故

统称为质量. 在国际单位制中，质量的单位是千克. 狭义相对论指出，速度大小为 v、静止质量为 m_0 的运动物体的质量为

$$m=\frac{m_0}{\sqrt{1-\left(\frac{v}{c}\right)^2}},$$

式中 c 为光速. 由此可推出物体的动能

$$E_k=mc^2-m_0c^2,$$

其中 mc^2 和 m_0c^2 分别为物体运动时的能量以及静止时的能量. 物理学上常用 E 和 E_0 表示该二者，即 $E=mc^2$，$E_0=m_0c^2$，称为质能关系式，表明物质的质量和能量存在一定的数量关系，是不可分割的. 工程上和日常生活中，物体运动速度较光速小得多，故质量为常量.

惯性质量(inertial mass)　见“惯性”.

惯性参考系(inertial reference system)　一种参考系. 是牛顿定律能在其中成立的参考系. 简称惯性系. 所有对惯性系匀速平动的参考系都是惯性系. 在实用中，一个参考系是否为惯性系，与人们考虑问题的精度有关. 例如，在日常生活和工程技术的多数问题中，地球自转引起地面物体的惯性离心力比其他作用力小得多，就忽略了这种影响，而把固连在地球上的坐标系当做惯性系. 而当问题需要考虑地球自转时，就采用地心为坐标原点，而坐标轴指向恒星的坐标系为惯性系.

非惯性参考系(non-inertial reference system)　一种参考系. 是牛顿运动定律在其中不能成立的参考系. 对惯性系有转动的参考系或加速平动的参考系都是非惯性参考系. 若它对惯性系以不变的加速度平动，就称为加速运动的参考系；若它对惯性系以不变角速度转动，就称为转动参考系.

万有引力定律(law of universal gravitation)　力学的基本定律之一. 它指出：具有质量的任意二物体间都存在着相互的吸引力，吸引力的大小为

$$F=G\frac{m_1m_2}{r^2}.$$

即该引力的大小与二者质量 m_1，m_2 的乘积成正比，与二者距离 r 的平方成反比，方向沿二者的连线，引力常数 G 约等于

$$6.6720\times10^{-11}\mathrm{m^3\cdot kg^{-1}\cdot s^{-2}}.$$

该定律由英国数学家、物理学家牛顿(Newton, I.)于 1687 年提出，是研究天体运行和航天技术的重要理论基础.

力的独立作用原理(principle of physical independence of force)　动力学原理. 指几个力同时作用于一个质点时，各力效应互不干扰的原理. 即它们共同产生的总加速度，等于各力单独产生的加速度的向量和. 是英国数学家、物理学家牛顿(Newton, I.)首先提出的，这样，就可把牛顿第二定律表示为 $m\boldsymbol{a}=\sum\boldsymbol{F}$，式中 m 为质量，$\boldsymbol{a}$ 为加速度，$\boldsymbol{F}$ 为作用力.

伽利略相对性原理(Galilei relativity principle)　亦称力学相对性原理. 力学原理. 即在任一惯性系内作任何的力学实验，都无法确定该惯性系相对其他惯性系是否运动的原理. 也可叙述为：牛顿力学的规律对任意惯性系都有同一的表达形式.

力学相对性原理(mechanical relativity principle)　即“伽利略相对性原理”.

惯性张量(inertia tensor)　亦称惯量张量. 动力学的基本概念之一. 指刚体相对一点质量分布的某些二次矩组成的二阶张量. 它表征刚体绕某点转动的惯性状况. 若记该点为 O，取直角坐标系 $Oxyz$，则称

$$I_{xx}=\sum m_i[y_i^2+z_i^2],$$
$$I_{yy}=\sum m_i[z_i^2+x_i^2],$$
$$I_{zz}=\sum m_i[x_i^2+y_i^2]$$

依次为刚体对 x 轴、y 轴和 z 轴的转动惯量(亦称为惯量矩、惯性矩、惯矩)；

$$I_{yz}=I_{zy}=\sum m_iy_iz_i,$$
$$I_{zx}=I_{xz}=\sum m_iz_ix_i,$$
$$I_{xy}=I_{yx}=\sum m_ix_iy_i$$

依次为刚体对 y 和 z，z 和 x，x 和 y 的惯性积(亦称为惯量积、离心距、离心转动惯量、惯积). 刚体对点 O 的惯性张量可表示为

$$[I]=\begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{yx} & I_{yy} & -I_{yz} \\ -I_{zx} & -I_{zy} & I_{zz} \end{bmatrix},$$

称这个矩阵为刚体在坐标系 $Oxyz$ 中的惯量矩阵. 刚体对过原点 O 的任一轴 l 的转动惯量为

$$I=[\alpha\ \beta\ \gamma][I][\alpha\ \beta\ \gamma]^{\mathrm{T}},$$

式中 α,β,γ 为 l 的方向余弦. 若以

$$\boldsymbol{\omega}=\omega_x\boldsymbol{i}+\omega_y\boldsymbol{j}+\omega_z\boldsymbol{k}$$

表示刚体的角速度，则它对点 O 的动量矩为

$$[G_x\ G_y\ G_z]=[I][\omega_x\ \omega_y\ \omega_z]^{\mathrm{T}}.$$

刚体的动能为

$$\frac{1}{2}[\omega_x\ \omega_y\ \omega_z][I][\omega_x\ \omega_y\ \omega_z]^{\mathrm{T}},$$

其中上角符号 T 代表矩阵转置.

刚体对任一点都存在三个相互垂直的轴，使惯性积等于零，从而使惯性张量表示为对角形矩阵，称这样的轴为惯性主轴. 刚体对惯性主轴的转动惯量称为主转动惯量. 刚体对其质心的惯性主轴称为中心惯性主轴，对中心惯性主轴的转动惯量称为中心主转动惯量. 在国际单位制中，转动惯量和惯性积的单位均为千克・米^2.

转动惯量(moment of inertia)　刚体绕轴转动

惯性的度量. 见"惯性张量".

惯性矩(moment of inertia) 即"转动惯量".

惯性积(product of inertia) 见"惯性张量".

离心转动惯量(centrifugal moment of inertia) 即"惯性积".

惯量半径(radius of inertia) 亦称回转半径. 动力学的基本概念之一. 指刚体相对一轴线质量分布的某种距离. 其平方乘以刚体质量就等于刚体对该轴的转动惯量.

回转半径(radius of inertia) 即"惯量半径".

惯性主轴(principal axis of inertia) 亦称惯量主轴. 动力学的基本概念之一. 指使刚体惯量积都等于零的直角坐标轴. 刚体相对任一点都存在这样的直角坐标,在这种坐标下,刚体的惯量张量可表示为对角形矩阵.

惯量椭球(ellipsoid of inertia) 刚体相对一点各方向转动惯量分布状况的图形,是以该点为中心,以三个惯性主轴为轴线的椭球. 若记该点为 O,取直角坐标系 $Oxyz$, I_{xx}, I_{yy}, I_{zz} 依次为刚体相对 x,y,z 的转动惯量, I_{xy}, I_{yz}, I_{zx} 依次为刚体相对 x 与 y, y 与 z, z 与 x 的惯性积,

$$I_{xx}x^2 + I_{yy}y^2 + I_{zz}z^2 - 2I_{xy}xy - 2I_{yz}yz - 2I_{zx}zx = k^2$$

即惯量椭球的方程,式中 k 为常数,一般常取为 1. 从点 O 到椭球面任一点 M 的距离的平方与刚体相对轴线 OM 的转动惯量 I 成反比,即 $|OM|^2 I = k^2$. 若令三个坐标轴和刚体惯性主轴重合,则椭球方程化为

$$I_{xx}x^2 + I_{yy}y^2 + I_{zz}z^2 = k^2.$$

若椭球相对某一惯性主轴是旋转椭球,则称这个轴为极轴,其垂直的轴称为赤道轴. 称刚体对极轴的转动惯量为极转动惯量,对赤道轴的转动惯量为赤道转动惯量. 常称这种半轴与主转动惯量平方根成反比的惯量椭球为潘索椭球和潘索惯量椭球. 若令椭球半轴的平方与主转动惯量成正比,即

$$\frac{\xi^2}{I_{xx}} + \frac{\eta^2}{I_{yy}} + \frac{\zeta^2}{I_{zz}} = \frac{R^4}{k^2},$$

则称为倒易惯量椭球,R 为任意常数. 过潘索椭球上任一点 M 作切平面,再由中心 O 作该平面的垂线,P 为垂足,则满足 $|\overrightarrow{OQ}| = R^2/|\overrightarrow{OP}|$ 的点 Q 就形成倒易惯量椭球上的动点. R 是点 O 与点 M 的距离. 若针对这种倒易惯量椭球作同样方式的变换,又可得到潘索椭球. 这两种椭球是互相倒易的.

潘索椭球(Poinsot ellipsoid) 见"惯量椭球".

马赫原理(Mach principle) 关于加速度相对性的原理:物体可具有相对遥远星系的加速度,不存在绝对加速度. 由奥地利学者马赫(Mach,E.)于 1872 年提出. 更正了牛顿绝对空间和绝对加速度的提法. 是对运动和空间认识的进一步深化,对物理学的发展有重要意义.

落体运动(motion of falling body) 物体的一类运动. 是物体由高处下落于地面的运动. 通常指下落高度远小于地球半径的情况. 落体除受重力外,还受空气等介质的阻力,而空气阻力在速度小时,与速度的一次方成正比;在 300m/s 附近,则与速度的二次方成正比;速度大时,则依赖于它的高次方. 一般表示为阻力 $R = \alpha\rho F\varphi(v)$,$\alpha$ 是物体形状系数,ρ 为介质密度,F 为落体居中截面(垂直于速度方向平面上的投影). 若 α 和 F 都小,下落高度不大,则可忽略空气阻力,重力加速度也不随高度而变,这种落体运动称为自由落体运动. 对由静止开始的自由落体运动,有公式

$$h = \frac{1}{2}gt^2, \quad v = gt, \quad v^2 = 2gh,$$

式中 h,g,t,v 分别为高度、重力加速度、下落经历的时间、落体速度. 对该问题值得提起的是历史上发生过的认识过程. 古希腊哲学家亚里士多德(Aristotle)曾提出,即使没有空气阻力,落体也必是重者快轻者慢,并为当时普遍接受. 将近两千年之后,意大利科学家伽利略(Galilei,G.)于 1604 年在实验基础上指出,忽略空气等介质的影响,落体运动是匀加速运动,与物体重量无关. 伽利略的学生在辩论中,亦巧妙地反问:如果重物比轻物下落快,那么一重一轻拴在一起又该如何? 使坚持亚里士多德观点的门徒陷入了困境. 落体运动问题的争论,是力学发展中的一件大事,伽利略不仅澄清了落体的本来面目,还首先提出了加速度概念. 他严格从实际出发和严密的逻辑推理的研究风格,对力学甚至整个科学的发展都是很有意义的.

自由落体运动(motion of free falling body)见"落体运动".

动力学基本定理(basic theorem of dynamics) 物体机械运动特征量与所受作用量关系的表述. 主要指动量定理、动量矩定理、动能定理. 曾为前人独立发现,但可由牛顿定律导出,在理论力学中以定理形式出现,使理论体系完美. 它们从整体上描述系统的动力学关系,物理意义明显,便于求解问题. 物体的一般机械运动可分解为随质心平动和绕质心转动,动量定理就说明其平动部分的动力学关系,动量矩定理说明其转动部分,而动能定理则包括了两个方面. 动能定理以标量形式出现,其他二者是向量关系.

动量定理(theorem of momentum) 物体动量与所受冲量关系的表述. 其微分形式为:质点系动量的微分等于所受外力元冲量的向量和,即

$$\mathrm{d}\boldsymbol{k} = \sum \boldsymbol{F}_i^{(e)}\mathrm{d}t,$$

或质点系动量对时间的导数等于所受外力主向量，即

$$\frac{\mathrm{d}\boldsymbol{k}}{\mathrm{d}t}=\sum \boldsymbol{F}_i^{(e)}.$$

以上 $\boldsymbol{k}=\sum m_i\boldsymbol{v}_i$ 为质点系的动量，$\boldsymbol{F}_i^{(e)}$ 为第 i 个质点所受外力，$\boldsymbol{F}_i^{(e)}\mathrm{d}t$ 为相应的元冲量，$\sum\boldsymbol{F}_i^{(e)}$ 为外力主向量. 其积分形式为：在某一时间间隔内，质点系动量的改变量，等于所受外力冲量的向量和，即

$$\boldsymbol{k}-\boldsymbol{k}_0=\sum\int_0^t \boldsymbol{F}_i^{(e)}\mathrm{d}t,$$

式中

$$\int_0^t \boldsymbol{F}_i^{(e)}\mathrm{d}t$$

为外力 $\boldsymbol{F}_i^{(e)}$ 在时间间隔 0 到 t 的冲量. 若质点系只有一个质点，以上就化为对一个质点的动量定理. 若质点系所受外力主向量为零，质点系的动量守恒；若外力主向量在某一轴线投影为零，则质点系的动量在该轴的投影量守恒. 此后两点即为动量守恒定律.

动量(momentum)　物体的平动特征量. 质点的动量为其质量与速度的乘积 $m\boldsymbol{v}$，质点系的动量为各个质点动量的向量和 $\sum m_i\boldsymbol{v}_i$，也等于总质量与质心速度的乘积 $M\boldsymbol{v}_c$. 动量是向量，在国际单位制中为千克・米/秒，或克・厘米/秒. 若质点速度 $\boldsymbol{v}$ 不是远小于光速的，则动量为

$$m_0\frac{\boldsymbol{v}}{\sqrt{1-\left(\frac{|\boldsymbol{v}|}{c}\right)^2}},$$

m_0，c 分别为质点的静止质量、光速.

冲量(impulse)　动力学的基本概念之一. 指力对物体作用的时间积累. 力 $\boldsymbol{F}$ 的元冲量为 $\mathrm{d}\boldsymbol{S}=\boldsymbol{F}\mathrm{d}t$，$\mathrm{d}t$ 为时间微分. 元冲量的积分就是力在相应时间间隔内的冲量

$$\boldsymbol{S}=\int_t^{t+\Delta t}\boldsymbol{F}\mathrm{d}t.$$

在国际单位制下，冲量的单位为千克・米/秒，或克・厘米/秒.

动量守恒(conservation of momentum)　见“动量定理”.

质心运动定理(theorem of motion of mass centre)　关于质点系运动的定理. 即质点系的总质量与质心加速度的乘积等于所受外力主向量. 若外力主向量为零，则质心速度不变；若外力主向量在某轴投影为零，则质心速度在该轴投影不变. 可见内力不影响质心运动.

质心(mass centre)　物体质量分布的特征量. 它是质量中心的简称. 设质点系各质点的质量为 m_i，对某点 O，向径为 $\boldsymbol{r}_i(i=1,2,3,\cdots)$，则质心 c 的向径为

$$\boldsymbol{r}_c=\frac{\sum m_i\boldsymbol{r}_i}{\sum m_i}.$$

质心相对质点系的位置只与质点系本身的质量分布有关，与 O 点位置或所选坐标系无关. 若取直角坐标系 $Oxyz$，则得质心的坐标表达式为

$$x_c=\frac{\sum m_ix_i}{\sum m_i},\ y_c=\frac{\sum m_iy_i}{\sum m_i},\ z_c=\frac{\sum m_iz_i}{\sum m_i}.$$

质心的位置关系到物体的稳定和机器的安全使用，工程上许多情况对物体质心位置有严格的规定.

动量矩(moment of momentum)　亦称角动量. 转动状态的特征量. 是物体动量对一点或一轴之矩. 动量矩相对一点为轴向量(参见“角速度”)，相对一轴为标量. 设对点 O 的动量矩为 $\boldsymbol{G}_O$，对轴 ζ 的动量矩为 G_ζ，当 O 在 ζ 上，必有 $(\boldsymbol{G}_O)_\zeta=G_\zeta$. 对于质点，$\boldsymbol{G}_O=\boldsymbol{r}\times m\boldsymbol{v}$，$G_\zeta=\pm mv_\perp h$，式中 m 为质量，$\boldsymbol{r}$ 为 O 到质点的向径，$\boldsymbol{v}$ 为速度，$v_\perp$ 为 $\boldsymbol{v}$ 对 ζ 的垂直分量，h 为 $\boldsymbol{v}$ 与 ζ 间的距离. 对于质点系，$\boldsymbol{G}_O=\sum\boldsymbol{G}_O^i$，$G_\zeta=\sum G_\zeta^i$，$i$ 表示质点系中一个质点的量. 可推证

$$\boldsymbol{G}_O=\boldsymbol{r}_c\times M\boldsymbol{v}_c+\boldsymbol{G}_c',$$

$\boldsymbol{r}_c$ 为 O 到质心 c 的向径，$\boldsymbol{v}_c$ 为 c 的速度，M 为总质量，$\boldsymbol{G}_c'$ 为质点系相对 c 上平动坐标运动引起的对 c 动量矩；同一质点系对不同点 A，B，有

$$\boldsymbol{G}_A=\boldsymbol{G}_B+\overrightarrow{AB}\times M\boldsymbol{v}_c.$$

$\boldsymbol{G}_A$，$\boldsymbol{G}_B$ 分别为质点系对 A，B 的动量矩. 对于刚体平动，$\boldsymbol{G}_O=\boldsymbol{r}_c\times M\boldsymbol{v}_c$；若刚体绕 O 转动，则

$$\boldsymbol{G}_O=\boldsymbol{\omega}\sum m_i\boldsymbol{r}_i^2-\sum m_i\boldsymbol{r}_i(\boldsymbol{r}_i\cdot\boldsymbol{\omega}),$$

或取直角坐标系 $Oxyz$，则

$$[G_1\ G_2\ G_3]^{\mathrm{T}}=[J][\omega_1\ \omega_2\ \omega_3]^{\mathrm{T}},$$

式中 G_1，G_2，G_3，ω_1，ω_2，ω_3 分别为 $\boldsymbol{G}_O$ 和角速度 $\boldsymbol{\omega}$ 在 x，y，z 的投影，$[J]$为矩阵表示刚体对 O 的惯量张量，T 表示转置；若刚体绕轴 ζ 转动，则 $G_\zeta=J_{\zeta\zeta}\omega$，$J_{\zeta\zeta}$，$\omega$ 分别为刚体对 ζ 的转动惯量、角速度；刚体做一般运动，有

$$\boldsymbol{G}_O=\boldsymbol{r}_c\times M\boldsymbol{v}_c+\boldsymbol{\omega}\sum m_i\boldsymbol{\rho}_i^2-\sum m_i\boldsymbol{\rho}_i(\boldsymbol{\rho}_i\cdot\boldsymbol{\omega}),$$

$\boldsymbol{\rho}_i$ 为质心 c 到刚体各点的向径.

角动量(angular momentum)　即“动量矩”.

冲量矩(moment of impulse)　动力学的基本概念之一. 指力矩对物体作用的时间积累. 设力 $\boldsymbol{F}$ 对点 P 的力矩为 $\boldsymbol{m}_P$，则 $\boldsymbol{m}_P\mathrm{d}t$ 为 $\boldsymbol{F}$ 对 P 的元冲量矩，也等于力的元冲量 $\boldsymbol{F}\mathrm{d}t$ 对点 P 的矩. 称

$$\int_0^t\boldsymbol{m}_P\mathrm{d}t$$

为力 $\boldsymbol{F}$ 在 0 到 t 时间间隔内对点 P 的冲量矩. 在国际单位制中，冲量矩单位为千克・米²/秒.

动量矩定理(theorem of moment of momentum)　物体转动特征量的变化与受力关系的定

理.其微分形式为:质点系对某定点(定轴)动量矩对时间的导数,等于所受外力对该点(轴)力矩的向量(代数)和.积分形式为:质点系对某定点(定轴)的动量矩在一段时间间隔内的改变量,等于所受外力在同一时间间隔内对该点(轴)冲量矩的向量(代数)和.当质点系只有一个质点时,该定理就成为对一个质点的情形.当质点系所受外力对某定点(定轴)的力矩向量(代数)和为零,则对该点(轴)的动量矩守恒,这就是动量矩守恒定律.质点系对于质心的动量矩定理为:质点系相对质心平动坐标系运动,引起的相对质心动量矩(对时间)的导数,等于所受外力对质心力矩的向量和.

动量矩守恒(conservation of moment of momentum) 见"动量矩定理".

回转力矩(gyroscopic moment) 亦称陀螺力矩.动力学的基本概念之一.指转子绕其轴转动时维持转轴方向不变的惯性力矩.它和转动惯量与角速度的乘积成正比.当外部物体带动转轴改变方向时,就产生一个反力矩,作用在迫使它改变方向的物体上,这种反力矩就是回转力矩:$\boldsymbol{M}=J\,\boldsymbol{\omega}\times\boldsymbol{\Omega}$,式中 J 和 $\boldsymbol{\omega}$ 分别是转子绕轴的转动惯量和角速度,$\boldsymbol{\Omega}$ 是迫使转轴改变方向的转动角速度.称这种现象为回转效应,或陀螺效应.例如,自行车前进中,车轮的 $\boldsymbol{\omega}$ 相对骑车人为由右向左,若施力矩让车向左转,则 $\boldsymbol{\Omega}$ 方向向上,骑车人右手必受反力矩 $\boldsymbol{M}$,其方向向前,对人体效应是使人向右倾不致倒于左侧,J 为车前轮绕转轴的转动惯量.实践中,搬动转动的机器、轮船和飞机的飞行,都有可能因转动体对载体产生回转力矩而受损害,所以应使 $\boldsymbol{\Omega}$ 小一些而减轻损害.只有当载体平移而不转动,才能使这种力矩为零.

陀螺力矩(gyroscopic moment) 即"回转力矩".

赖柴尔定理(Résal theorem) 动量矩定理的几何诠释:质点系对定点 O 的动量矩向量的端点速度,等于所受外力系对 O 的主矩.若令 $\boldsymbol{H}$ 表示质点系对点 O 的动量矩,$\boldsymbol{L}$ 为所受外力系对 O 的主矩,则有 $\mathrm{d}\boldsymbol{H}/\mathrm{d}t=\boldsymbol{L}$,等式左边的 $\mathrm{d}\,\boldsymbol{H}/\mathrm{d}t$ 即为矢量 $\boldsymbol{H}$ 端点的速度.这种诠释为法国工程师赖柴尔(Résal, H.)给出.

变质量质点(particle with variable mass) 一种力学模型.指质量为时间函数的质点.平动物体的质量不断并入或分出而不改变质心相对位置就可视为此类质点,例如飞行中的火箭、收割机后面接收农产品的汽车.其运动微分方程为

$$m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=\boldsymbol{F}+\boldsymbol{v}_{\mathrm{r}}\frac{\mathrm{d}m}{\mathrm{d}t}.$$

由密歇尔斯基(Мещерский, И. В.)于 1897 年给出,称为密歇尔斯基方程.式中 $m=m(t)$ 是质点的质量,$\boldsymbol{F}$ 为质点并入或分出的微量上全部外力向量之和,$\boldsymbol{v}$ 是质点的速度,$\boldsymbol{v}_{\mathrm{r}}$ 是即将并入或分出的微量相对该质点的速度.该模型不是考虑速度接近光速时质量依赖于速度而变化,只是考虑质量的不断加入和分出.该模型忽略了物体的转动以及质心相对位置的变化,不同于一般的变质量体.从一般意义上考虑,变质量体的动力学问题比较复杂,超出了经典力学的范畴.

密歇尔斯基方程(Meshchersky equation) 见"变质量质点".

齐奥尔柯夫斯基公式(Tsiolkovsky formula) 火箭铅直升空速度表达式.是俄国科学家齐奥尔柯夫斯基(Циолковский, К. Э.)于 1914 年提出的.其一为不计外力只考虑喷气反推力的情形,其二为加入重力的情形.即

$$v=v_0+v_2\ln\frac{M_0}{M},$$

$$v=v_0-gt+v_2\ln\frac{M_0}{M},$$

式中 v, v_0 以及 v_2 分别为火箭瞬时速度、原始速度以及喷放速度(喷气相对火箭的速度)的绝对值,M 为火箭瞬时质量,M_0 为原始总质量(包括所带燃料质量 M_2 和火箭躯体质量 M_1),g 为重力加速度.这两个表达式所表述的问题又称为齐奥尔柯夫斯基第一、第二问题.

在上述第一公式中,若令 $v_0=0$,且记 v_1 为燃料用完时火箭的最大速度,则

$$\frac{v_1}{v_2}=\ln\frac{M_0}{M_1}\ 或\ \frac{v_1}{v_2}=\ln\left(1+\frac{M_2}{M_1}\right).$$

称 M_2/M_1 为齐奥尔柯夫斯基数,v_1/v_2 为火箭的特征速度.若 M_0/M_1 扩大为原来的 n 次方,则v_1/v_2只扩大为原来的 n 倍.可见,扩大 M_0/M_1 不是提高 v_1 的好方法,不如提高 v_2,即加大燃料不如提高喷放速度,但喷放速度也受技术条件限制.齐氏提出多级火箭,正是近代实践所循之法.

齐奥尔柯夫斯基数(Tsiolkovsky number) 见"齐奥尔柯夫斯基公式".

火箭特征速度(characteristic velocity of rocket) 见"齐奥尔柯夫斯基公式".

功(work) 力学的基本概念之一.指力对物体作用沿路径的积累.设力 $\boldsymbol{F}$ 在物体上作用点的矢径为 $\boldsymbol{r}$,则力所做的元功为 $\mathrm{d}W=\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}$,当做用点由 A 到达 B,则力所做的功为

$$W=\int_A^B\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}.$$

$\boldsymbol{F}$ 也可以理解为同一点所受几个力的合力.在国际单位制中,功的单位为焦耳(=牛顿·米),在工程单位制中,为千克力·米,1kgf·m=9.8J.

关于功概念的称谓，法国力学家、数学家卡诺(Carnot，N. L. S.)曾在1824年使用“动力效应”、“机械能力”的名称，在1826—1829年，法国数学家、力学家彭赛列(Poncelet，J.-V.)等人在论文中使用了“功”这个词，以后就这样沿用下来.

功率(power)　力学单位. 指单位时间里力作的功. 可表示为

$$P=\frac{\mathrm{d}W}{\mathrm{d}t},$$

W 为功，t 为时间. 若力 $\boldsymbol{F}$ 的作用点速度为 $\boldsymbol{v}$，则 $P=\boldsymbol{F}\cdot\boldsymbol{v}$；若力作用于定轴转动刚体上，则 $P=\boldsymbol{M}\cdot\boldsymbol{\omega}$，$\boldsymbol{M}$ 为力对转轴的力矩，$\boldsymbol{\omega}$ 为刚体角速度. 在国际单位制中，功率的单位为瓦特(＝焦耳/秒)，在工程单位制中为千克力·米/秒，1kgf·m/s＝9.8W；实用上还常用公制马力(ps)及英制马力(hp)，1ps＝735.5W，1hp＝745.7W.

动能(kinetic energy)　物体机械运动的一种特征量. 它度量物体由于机械运动而具有的作功本领，是非负标量. 若记动能为 T，对质点，有

$$T=\frac{1}{2}mv^2,$$

m，v 分别为其质量和速度. 对质点系，有

$$T=\sum\frac{1}{2}m_iv_i^2$$

(即系内各质点动能总和)，或由柯尼希定理：

$$T=\frac{1}{2}Mv_c^2+\sum\frac{1}{2}m_iv_i'^2,$$

M 为质点系总质量，v_c 为质心 c 的速度，v_i' 为质点相对 c 的平动坐标的速度. 对刚体的平动，有

$$T=\frac{1}{2}Mv^2;$$

绕定轴转动，有

$$T=\frac{1}{2}J\omega^2,$$

J 为转动惯量，ω 为角速度；平面运动，有

$$T=\frac{1}{2}Mv_c^2+\frac{1}{2}J_c\omega^2,$$

J_c 为刚体对垂直平面且过质心之轴的转动惯量；绕定点 O 转动，有

$$T=\frac{1}{2}\boldsymbol{\omega}\cdot\boldsymbol{J}\cdot\boldsymbol{\omega}=\frac{1}{2}\boldsymbol{\omega}\cdot\boldsymbol{G},$$

$\boldsymbol{\omega}$ 为向量表示的刚体角速度，$\boldsymbol{J}$ 为对 O 的惯量张量，$\boldsymbol{G}$ 为对 O 的动量矩. 若取直角坐标系 $Oxyz$，则有

$$T=\frac{1}{2}[\omega_1\ \omega_2\ \omega_3][J][\omega_1\ \omega_2\ \omega_3]^{\mathrm{T}},$$

$[J]$ 为矩阵表示的惯量张量，右上角 T 表示转置. 自由刚体运动，有

$$T=\frac{1}{2}Mv_c^2+\frac{1}{2}\boldsymbol{\omega}\cdot\boldsymbol{J}_c\cdot\boldsymbol{\omega},$$

$\boldsymbol{J}_c$ 为刚体对质心的惯量张量，在国际单位制中，动能的单位为焦耳(＝牛顿·米)，在厘米·克·秒制中为尔格(＝达因·厘米). 动能在计量物体机械运动时，既包括物体的平动部分，也包括了转动部分. 而动量只包含其平动部分. 动能这种量既能计量机械运动之间的传递，也能计量机械运动和其他运动之间的传递和转换.

活力(active force)　动能的别称. 德国数学家、哲学家莱布尼茨(Leibnitz，G. W.)于1684年提出时，是指动能的2倍，即对于质点，动能为 mv^2 (m 为质点质量，v 为速度). 原意是与静止物体的“死力”如压力作对照. 蒸汽机的发明，实现了机械运动和热的转换，在此基础上，后人在研究物体机械运动量的传递以及这种量和其他形式的运动量转换中，把“活力”和现今通用的“动能”等同起来，即对质点的活力为

$$\frac{1}{2}mv^2.$$

活力这个词现已被动能取代，很少使用.

柯尼希定理(Koenig theorem)　见“动能”.

动能定理(theorem of kinetic energy)　物体动能变化和所受力作功的关系定理. 其微分形式为：质点系动能 T 的微分，等于所受力的元功 δW 之和，即

$$\mathrm{d}T=\sum\delta W;$$

积分形式为：在任意有限的路程中，质点系动能的改变量等于所受力作功之和. 在理想约束(约束力作功之和为零)之下，质点系动能的微分等于所受主动力的元功之和；在理想约束下，在有限的路程中，质点系动能的改变量等于所受主动力作的功. 该定理的积分形式能给出系统的广义速度(多为线速度和角速度)、位置和力的关系；其微分形式能对一个自由的系统给出加速度或角加速度和受力的关系. 该定理为求解问题的方便工具.

认识该定理曾有一段漫长而曲折的过程，并随着时间的推移扩展为更一般的原理. 起初，意大利科学家伽利略(Galilei，G.)给出了真空中落体距离和速度的关系；后来，荷兰数学家、物理学家惠更斯(Huygens，C.)针对摆的振动得出了“摆到平衡位置的距离相等时，其速度平方必相等”的结论. 这是沿着观察、实验上升为理论的路径迈向一般原理的进程. 牛顿三大运动定律的出现(1687)，提供了另一条认识该定理的路径. 瑞士数学家约翰第三·伯努利(Bernoulli，Johann Ⅲ)和丹尼尔第一·伯努利(Bernoulli，Daniel Ⅰ)由牛顿定律导出了和该定理非常接近的“活力”定理，基本上使该定理确立于已有的原理之上. 当然，该定理只考虑不同形式的机械

运动之间互相传递的数量关系,不涉及其他运动形态.随着生产和科学技术的发展,其他物理现象的研究逐步深入,尤其是蒸汽机的出现、热的概念的产生,客观上需要把能直接观察到的机械运动与不易直接观察到的“隐蔽”形态的运动之间建立起数量关系,要求这部分最先形成的力学理论向其他科学领域开拓.1824年,法国力学家、数学家卡诺(Carnot,N. L. S.)和1841年,莫尔(Mohr,O.)先后提出连同功和活力在内,热是表征运动的又一种量.1847年,亥姆霍兹(Helmholtz,H. L. F.)指出活力和弹性力所做的功之和是常量,并指出在弹性体的运动中计入热能的必要性,这就等于把物体机械运动的能和分子运动的能统一计量.1853年,兰金(Rankin)强调:任何能的形态都可由某种功的变换而得到.从而把不同运动形式的能的传递和转换关系明朗化了,成为目前公认的能量原理(各种形式的能量可通过作功或传热等方式互相转换,总能量不变)的完备形式.这里的动能定理虽只属于能量原理的一个特殊情况,但在原理认识的过程中,曾起过开路的作用.

保守力场(conservative field)　亦称势力场,或有势力场.力场的一种.质点在某空间的每一位置都受到大小、方向完全确定的力,称这空间为力场.若在力场内,力所作的功和路径无关,只与力的作用点的起点和终点位置有关,则称这力场为保守力场.重力、弹性力和万有引力形成的力场都是保守力场.保守力场内存在势函数 $U(x,y,z)$,使质点所受的力在 x 轴、y 轴、z 轴上的投影分别为

$$\frac{\partial U}{\partial x},\quad \frac{\partial U}{\partial y},\quad \frac{\partial U}{\partial z}.$$

有非保守力的力场,不是保守力场.

势力场(potential field)　即“保守力场”.

有势力场(potential field)　即“保守力场”.

势函数(potential function)　见“保守力场”.

势能(potential energy)　亦称位能.力学的基本概念之一.指物体在势力场中由所处位置具有的作功本领.是相对的量.物体在甲位置相对乙位置的势能是由甲位置到达乙位置时所受势力作的功;若取某位置为零势能位置,那么物体在任一位置的势能,等于由该位置到达零势能位置时所受势力做的功.例如,在地球表面附近,重量为 P、高度为 H 的物体相对地面的势能为 PH.若令势函数在所选的零势能位置为零,则质点在势力场中任一点的势能函数,就等于该点的势函数冠以负号,即势能函数 $V=-U$,U 为势函数.在国际单位制中,势能的单位为焦耳(=牛顿·米).

势能函数(potential energy function)　势力场中势能依赖于位置的函数,一般记为 $V(x,y,z)$.若 $U(x,y,z)$ 为势函数,则

$$V(x,y,z)=U(x_0,y_0,z_0)-U(x,y,z),$$

其中 (x_0,y_0,z_0) 为零势能位置点的坐标.

机械能(mechanical energy)　物体的动能、势能之和.

机械能守恒(conservation of mechanical energy)　力学原理.指物体的机械能总量不变.物体在保守力场中运动不受其他非保守力时,动能、势能可相互转换,但总量不变.例如,重力场就是保守力场,在这个场中,若不计摩擦力,物体的机械能就守恒.对万有引力场也是这样.

惯性力(inertial force)　一种虚拟的力.它等于质点的质量乘以加速度,再取负号(假想它施于该质点本身).虽在非惯性系能被感知,具有动力学效应,但因不存在施力者,不是物体之间的相互作用,故为非真实的力.借助这个概念,根据达朗贝尔原理建立的动静法,广泛地应用于力学和工程领域,而建立的相对非惯性系动力学方程,则可把惯性系中的某些方法平行地引入非惯性系.和质量相乘的加速度,在动静法中是相对惯性系的整个加速度,在相对非惯性系的动力方程中是牵连加速度和科里奥利加速度,相应地称为惯性力、牵连惯性力和科里奥利惯性力.

惯性力概念,最早由德国科学家开普勒(Kepler,J.)提出,是指物体对迫使它改变运动状态者呈现的反抗力,即动态反作用力,属于1687年英国数学家、物理学家牛顿(Newton,J.)指出的作用力和反作用力原理中反作用力的一部分.1743年,法国数学家、力学家达朗贝尔(d'Alembert,J. le R.)提出了后人以他的名字命名的原理,指出质点的质量乘以加速度取负号与所受的主动力、约束力满足静力学平衡条件,后人归纳为质点的惯性力和所受真实的力相平衡,引出了动静法.在动力学普遍方程及分析力学中,都从这个观点出发,结合其他原理,建立了完美的理论体系,为解决实践中提出的力学问题提供了强有力的工具.

惯性力是虚拟的力还是真实的力,学术界也有争议.“虚拟说”强调它没有施力者,也就不存在反作用力;“真实论”强调它在非惯性系中具有力的动态和静态效应.总之,双方都承认惯性力所具有的全部特征,只需在力的定义方面再加斟酌.此应属近代物理学范围.例如,考虑广义相对论,已超出牛顿力学的领域.在人们所研究的范围内,既已出言“力是物体间的相互作用”,逻辑推人们于“虚拟说”的一边.

达朗贝尔原理(d'Alembert principle)　非自由质点系动力学的普遍原理:质点系各质点的惯性力和所受的外力(含主动力和约束力)在任一瞬时满足平衡条件,即它们的主向量为零:

$$\sum \boldsymbol{F} + \sum \boldsymbol{g} = 0;$$

对任一点的主矩为零：

$$\sum \boldsymbol{L}_P + \sum \boldsymbol{M}_P = 0.$$

式中 $\boldsymbol{F}$,$\boldsymbol{g}$ 分别为外力、惯性力,$\boldsymbol{L}_P$,$\boldsymbol{M}_P$ 依次为该两种力对任一点 P 之矩. 当质点系中只有一个质点时,则可说任一瞬时质点的惯性力和外力满足平衡条件. 该原理由法国数学家、力学家达朗贝尔(d'Alembert,J. le R.)于1743年在《动力学论文》中提出,认为非自由质点系在任一瞬时的损失力与约束力平衡. 而损失力即所受主动力和每一质点的 $-m_i\boldsymbol{a}$(质量乘以加速度,再取负号,人们称为惯性力). 换一说法,即质点系所受的主动力、约束力和其惯性力主向量为零,并对任一点主矩为零. 而整个内力成对出现,本身就满足这两个条件. 基于该原理发展的动静法,在工程力学中得到了广泛的应用,它和虚位移原理结合,又建立了动力学普遍方程,成为分析动力学的基础. 对法国数学家、力学家拉格朗日(Lagrange,J.-L.)和哈密顿力学体系的形成,起了重要作用.

动静法(method of kineto-statics)　动力学问题用静力学手段处理的一种方法. 引入了惯性力概念,根据达朗贝尔原理就直接产生这种方法. 因具有直观和方便的特点,故广泛应用于力学和工程领域. 对于自由质点,该方法只相当于牛顿方程进行一次移项,形式上属于简单的数学处理,但是对于较复杂的受约束质点系动力学问题,优点就显露出来. 例如,对分析动约束力以及动载荷下的内力分布,就很明显.

向心力(centripetal force)　一种特殊的力. 是曲线运动中质点所受指向该处的曲率中心的力. 大小为 mv^2/r,r 为该处的曲率半径,m,v 分别为质点的质量、速率. 施力者是迫使质点进行曲线运动的物体. 例如,由高处水平抛出的石子,略去空气阻力,轨迹为抛物线,刚出手时向心力为重力,运行中向心力为重力在瞬时位置主法向的分力(重力的切向分力提供质点的切向加速度).

离心力(centrifugal force)　一种特殊的力. 是向心力的反作用力. 质点施于迫使它进行曲线运动的物体上力的大小. 力的大小等于 mv^2/r,力的方向沿主法线而背离曲率中心. 另外,也常作为惯性离心力的简称,在相对惯性系有角速度 ω 的参考系中,质点由 ω 引起的牵连惯性力,称为惯性离心力,简称离心力. 例如,汽车转弯时,车座施于乘客之力就有一部分为向心力,它迫使乘客随车而转,乘客给车座相应的反作用力就是离心力. 然而,按照惯性力的概念,认为乘客自身就受有惯性离心力,简称离心力,大小等于车座给乘客的向心力,方向与之相反. 因而乘客在汽车中"平衡",这正好符合人们在非惯性参考系中的直觉. 日常人们所说的甩干机通过离心力将衣服甩干,实际上就是用了惯性离心力的简称.

约束反力(reactive force of constraint)　约束的作用力. 指被约束体受约束体的作用力,简称约束力. 约束反力分动静两类,动约束力亦称附加动反力,和被约束体的加速运动有关. 例如,电梯以加速度 a 上升,其中质量为 m 的人受向上的动约束反力大小为 ma;绕轴转动的偏心转子要受到轴承的动约束反力,轴承必受到动压力. 静约束力大小取决于被约束物体所受的主动力,方向与被阻碍的运动方向相反,作用点为约束和被约束体的接触点. 例如,对于光滑接触面,约束力通过接触点垂直公切面指向被约束体;对于光滑球铰链,约束力通过铰心,方向不定,但可用三个未定的垂直分量表示. 在工程技术方面,约束力的分析计算十分重要. 而对于高速转动的转子,动约束力有时能达到静约束力的10倍以上,这种力的消除成为高速转动机械的重要技术问题. 理论上讲,使转轴通过转子的质量中心并成为惯性主轴就可使这种动约束反力为零.

动约束力(dynamic constraint force)　见"约束力".

附加动反力(dynamic constraint reaction force)　动约束力的别称. 见"约束反力".

中心力(central force)　亦称有心力或辏力. 一种特殊的力. 运动质点所受力的作用线总是通过一个称为力心的固定点,这种力就称为中心力. 该质点对力心动量矩守恒. 运动轨迹和力心在同一平面,质点向径所扫面积对时间的变化率——面积速度为常量. 称此为面积速度定律. 若以力心 O 为极点,则极坐标表示的质点轨迹方程为

$$mc^2\frac{1}{r^2}\left[\frac{\mathrm{d}^2}{\mathrm{d}\varphi^2}\left(\frac{1}{r}\right)+\frac{1}{r}\right]=-F_r,$$

式中 m 为质点的质量,c 为面积速度的2倍,F_r 为力的径向投影. 称这公式为比耐公式,可由力确定轨迹或由轨迹确定力,在天体力学中有很多应用.

面积速度定律(law of area velocity)　见"中心力".

比耐公式(Binet formula)　见"中心力".

开普勒定律(Kepler laws)　行星运行的基本定律. 它是德国科学家开普勒(Kepler,J.)提出的行星运行规律. 共三条:

1. 行星轨道是以太阳为一焦点的椭圆.

2. 由日心引到行星的向径扫过的面积速度为常量.

3. 行星运行周期的平方和轨道半长轴的立方成正比.

开普勒定律是开普勒根据丹麦天文学家第谷

(Tycho Brahe)观察行星运动的丰富资料总结出来的. 英国数学家、物理学家牛顿(Newton,I.)又从此定律出发得出万有引力定律. 反过来,以上三定律又可以从牛顿第二定律和万有引力定律推出.

落体东偏(east deviation of falling body) 动力学现象. 指科氏惯性力使落体偏离铅垂线的现象. 地球自转使其表面具有向心加速度$-|\boldsymbol{\omega}|^2\boldsymbol{\rho}$和科里奥利加速度$2\boldsymbol{\omega}\times\boldsymbol{v}_r$,式中$\boldsymbol{\omega}$为地球自转角速度,方向沿南北轴,自南至北,$\boldsymbol{\rho}$为垂直地球南北轴而引到物体的向径,$\boldsymbol{v}_r$为物体相对地球运动的速度. 这样,物体相对地球的运动就需计入牵连惯性力$m|\boldsymbol{\omega}|^2\boldsymbol{\rho}$和科氏惯性力$-2m\boldsymbol{\omega}\times\boldsymbol{v}_r$,$m$为物体质量. $m|\boldsymbol{\omega}|^2\boldsymbol{\rho}$和地心引力合成为物体的重力$m\boldsymbol{g}$($\boldsymbol{g}$为重力加速度,沿铅垂线指向地面),使其铅垂线不通过地心而和赤道平面夹角为φ. 取落体原来位置为原点,$\boldsymbol{r}$为其位置向径,不计空气阻力,则有

$$m\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}=m\boldsymbol{g}-2m\boldsymbol{\omega}\times\boldsymbol{v}_r.$$

若无初速度,则近似取$\boldsymbol{v}_r=\boldsymbol{g}t$,可积分得到

$$\boldsymbol{r}=\boldsymbol{g}\frac{t^2}{2}+|\boldsymbol{\omega}|g\cos\varphi\frac{t^3}{3}\boldsymbol{\tau},$$

$\boldsymbol{\tau}$为向东的单位向量. 此表明物体下落的同时,也向东运动. 物体若在赤道上空100m处下落,可有近2cm的向东偏离. 在北京地区同样高度下落,则向东偏离约1.67cm. 进一步考虑相对速度的变化实况,还将算得在北半球尚有一定的南偏,在南半球则有一定的北偏,不过数量很小. 若精确地考虑偏离的实际数量,还应计入月球引力的影响.

泊松方程(Poisson equations) 刚体运动学的重要方程. 它指出刚体绕定点运动中角速度和固定轴线的方向余弦之间的运动学关系:

$$\frac{\mathrm{d}\gamma_1}{\mathrm{d}t}=\gamma_2\omega_3-\gamma_3\omega_2,$$

$$\frac{\mathrm{d}\gamma_2}{\mathrm{d}t}=\gamma_3\omega_1-\gamma_1\omega_3,$$

$$\frac{\mathrm{d}\gamma_3}{\mathrm{d}t}=\gamma_1\omega_2-\gamma_2\omega_1,$$

式中$\omega_1,\omega_2,\omega_3$是角速度在运动坐标轴上的投影,该坐标系的三个轴重合于刚体对定点的惯量主轴,γ_1,γ_2,γ_3是固定坐标系的一个轴对运动坐标系的方向余弦,因法国数学家、力学家泊松(Poisson, S.-D.)推出而得名. 在重刚体绕定点运动的求解中,和相应的动力学方程联立,可使方程组封闭.

重刚体绕定点运动(motion of heavy rigid body about fixed point) 一种刚体运动. 指所受主动力只是重力的刚体绕定点运动. 此运动在刚体绕定点运动的研究中具有特殊地位,曾吸引许多数学、力学工作者寻求其解析解. 其中欧拉情形、拉格朗日情形、柯瓦列夫斯卡娅情形可在任意初始条件下获得足够的首次积分,从而最后求得解析解. 记定点为O,取固定于地球的坐标系为$O\xi\eta\zeta$,设$O\zeta$轴铅垂向上;并设随刚体运动的坐标系为$Oxyz$,其各轴重合于刚体对O的惯量主轴. J_1,J_2,J_3依次为刚体对x轴、y轴、z轴的转动惯量,x_c,y_c,z_c是刚体质心在动坐标系的坐标,$\omega_1,\omega_2,\omega_3$是角速度分别在$x$轴、$y$轴、$z$轴上的投影,$\gamma_1,\gamma_2,\gamma_3$分别是$O\zeta$在$x$轴、$y$轴、$z$轴的方向余弦,$m$为刚体质量,$g$为重力加速度,基本方程可表示为

$$J_1\frac{\mathrm{d}\omega_1}{\mathrm{d}t}+(J_3-J_2)\omega_2\omega_3=mg(\gamma_2z_c-\gamma_3y_c),$$

$$J_2\frac{\mathrm{d}\omega_2}{\mathrm{d}t}+(J_1-J_3)\omega_3\omega_1=mg(\gamma_3x_c-\gamma_1z_c),$$

$$J_3\frac{\mathrm{d}\omega_3}{\mathrm{d}t}+(J_2-J_1)\omega_1\omega_2=mg(\gamma_1y_c-\gamma_2x_c);$$

$$\frac{\mathrm{d}\gamma_1}{\mathrm{d}t}=\omega_3\gamma_2-\omega_2\gamma_3,$$

$$\frac{\mathrm{d}\gamma_2}{\mathrm{d}t}=\omega_1\gamma_3-\omega_3\gamma_1,$$

$$\frac{\mathrm{d}\gamma_3}{\mathrm{d}t}=\omega_2\gamma_1-\omega_1\gamma_2.$$

若能求得$\omega_1,\omega_2,\omega_3$,而$\gamma_1,\gamma_2,\gamma_3$为时间$t$的函数,就可由欧拉运动学方程

$$\omega_1=\frac{\mathrm{d}\psi}{\mathrm{d}t}\gamma_1+\frac{\mathrm{d}\theta}{\mathrm{d}t}\cos\varphi,$$

$$\omega_2=\frac{\mathrm{d}\psi}{\mathrm{d}t}\gamma_2-\frac{\mathrm{d}\theta}{\mathrm{d}t}\sin\varphi,$$

$$\omega_3=\frac{\mathrm{d}\psi}{\mathrm{d}t}\gamma_3+\frac{\mathrm{d}\varphi}{\mathrm{d}t}$$

确定欧拉角ψ,θ,φ随时间t变化的规律. 而机械能守恒:

$$\frac{1}{2}(J_1\omega_1^2+J_2\omega_2^2+J_3\omega_3^2)+mg(x_c\gamma_1+y_c\gamma_2+z_c\gamma_3)=\text{常数},$$

刚体相对$O\zeta$轴动量矩守恒:

$$J_1\omega_1\gamma_1+J_2\omega_2\gamma_2+J_3\omega_3\gamma_3=\text{常数},$$

几何关系$\gamma_1^2+\gamma_2^2+\gamma_3^2=1$是三个首次积分,对于这里的前六个方程,只需第四个首次积分,就足以求得诸ω和γ关于t的函数关系,然后可进一步由欧拉方程把欧拉角表示为t的函数. 而欧拉、拉格朗日和柯瓦列夫斯卡娅三种情形就是能找到第四个首次积分的情形.

欧拉情形(Euler condition) 亦称欧拉-潘索情形. 任意初条件下刚体绕定点运动可获得解析解的一种特殊情况. 这里是指刚体绕定点运动不受外力矩的情形. 即合外力通过固定点因惯性而运动的情形. 对于重力场,刚体的质心重合于固定点且无其他主动力即属此情形. 瑞士数学家欧拉(Euler,L.)给出解析解. 潘索(Poinsot,L.)给出几何解释. 这种情

形的动能 T 和对固定点的动量矩 $\boldsymbol{G}$ 都恒定，因而降低了动力学方程的阶数，可用椭圆函数表示其解. 设 O 为固定点，$O\xi\eta\zeta$ 为固定于惯性空间的直角坐标系，令 $O\zeta$ 和 $\boldsymbol{G}$ 重合，$Oxyz$ 为随刚体而动的直角坐标系，取其轴为刚体对 O 的惯量主轴，$J_1 \geqslant J_2 \geqslant J_3$ 为主转动惯量，ψ, θ, φ 依次为进动角、章动角和自转角，$\omega_1, \omega_2, \omega_3$ 依次为角速度向量 $\boldsymbol{\omega}$ 在 x, y, z 三坐标轴上的投影，t 为时间，t_0 为初时刻，α, γ, ψ_0 依次为 ω_1, ω_3, ψ 在 $t=t_0$ 的值，则

$$\omega_1 = \alpha \operatorname{cn} \sigma(t - t_0),$$
$$\omega_2 = \beta \operatorname{sn} \sigma(t - t_0),$$
$$\omega_3 = \gamma \operatorname{dn} \sigma(t - t_0);$$
$$\psi = |\boldsymbol{G}| \int_{t_0}^{t} \frac{2T - J_3\gamma^2 \operatorname{dn}^2 \sigma(t - t_0)}{|\boldsymbol{G}|^2 - J_3^2 \gamma^2 \operatorname{dn}^2 \sigma(t - t_0)} \mathrm{d}t + \psi_0,$$
$$\cos\theta = \frac{J_3 \gamma \operatorname{dn} \sigma(t - t_0)}{|\boldsymbol{G}|},$$
$$\tan\varphi = \sqrt{\frac{J_1(J_2 - J_3)}{J_2(J_1 - J_3)}} \frac{\operatorname{cn}\sigma(t - t_0)}{\operatorname{sn}\sigma(t - t_0)}$$

为解. 式中 $\operatorname{sn}\sigma = \sin\Omega$, $\operatorname{cn}\sigma = \cos\Omega$, $\operatorname{dn}\sigma = \sqrt{1-k^2\sin^2\Omega}$，由椭圆积分

$$\sigma = \int_0^{\Omega} \frac{\mathrm{d}v}{\sqrt{1 - k^2 \sin^2 v}}$$

确定它们的关系为

$$\sigma^2 = \gamma^2 \frac{(J_1 - J_3)(J_2 - J_3)}{J_1 J_2},$$
$$\beta^2 = \alpha^2 \frac{J_1(J_1 - J_3)}{J_2(J_2 - J_3)},$$
$$k^2 = \frac{\alpha^2 J_1(J_1 - J_2)}{\gamma^2 J_3(J_2 - J_3)}.$$

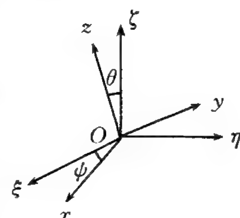
在运动中，惯量椭球(其方程为 $J_1x^2+J_2y^2+J_3z^2=1$)将随刚体运动，设 P 为瞬时轴(即角速度所在直线)和椭球的交点，称为极点；过 P 的切平面必在惯性空间固定，这平面到 O 点的距离为 $\sqrt{2T}/|\boldsymbol{G}|$，且与 $\boldsymbol{G}$ 正交，称为潘索平面. 椭球将无滑动地沿潘索平面滚动，并且以大小等于 $|\boldsymbol{\omega}|$ 的角速度绕 OP 转动；同时刚体的动瞬轴锥面以同一角速度沿定瞬轴锥面滚动.

若刚体对固定点 O 的惯量椭球为旋转对称的，不妨设 $J_1=J_2$，则这种运动就成为规则进动. 仍令固定坐标系 $O\xi\eta\zeta$ 的 ζ 轴和恒定的动量矩向量 $\boldsymbol{G}$ 同向，很易解得章动角 $\theta=\theta_0$，自转角 $\varphi=\omega_{自} t+\varphi_0$，进动角 $\psi=\omega_{进} t+\psi_0$. 式中 $\theta_0, \varphi_0, \psi_0$ 依次为 θ, φ, ψ 的起始值；$\omega_{自} = (\gamma - |\boldsymbol{G}|\cos\theta_0/J_1)$ 为自转角速度，$\omega_{进} = |\boldsymbol{G}|/J_1$ 为进动角速度.

欧拉-潘索情形(Euler-Poinsot condition)　即“欧拉情形”.

潘索平面(Poinsot plane)　见“欧拉情形”.

拉格朗日情形(Lagrange condition)　任意初条件下刚体绕定点运动可获解析解的一种情形. 这里是指对称回转器的情形. 即重刚体(所受主动力只是重力的刚体)对所绕固定点的惯量椭球旋转对称、质心位于对称轴的情形. 称该轴为动力对称轴. 法国数学家、力学家拉格朗日(Lagrange, J.-L.)和泊松(Poisson, S.-D.)曾先后对该问题进行了有效的研究，故亦称为拉格朗日-泊松情形. 若记固定点为 O，$O\xi\eta\zeta$ 与 $Oxyz$ 分别为固定于地球和随刚体运动的坐标系，$O\zeta$ 轴铅直向上，Oz 轴沿刚体的动力对称轴. 不失一般性，取 Ox 轴在 $O\xi\eta$ 平面内，ψ 为 $O\xi$ 轴到 Ox 轴的夹角(进动角)，θ 为 $O\zeta$ 到 Oz 的夹角(章动角). 刚体的运动分解为随着 $Oxyz$ 坐标系的进动和章动，加上绕 z 轴的自转，令 φ 为自转角. 在这里，刚体对 z 轴和 ζ 轴动量矩守恒，并且总机械能也守恒，记 mg 为刚体的重力，可得到以下三个动力学方程的首次积分：



$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} + \frac{\mathrm{d}\psi}{\mathrm{d}t}\cos\theta = \Omega \quad (常量),$$
$$J_1 \frac{\mathrm{d}\psi}{\mathrm{d}t}\sin^2\theta + J_3\Omega\cos\theta = H \quad (常量),$$
$$J_1\left[\left(\frac{\mathrm{d}\psi}{\mathrm{d}t}\right)^2 \sin^2\theta + \left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2\right] + 2mga\cos\theta = 2E - J_3\Omega^2,$$

常量 E 为刚体的总机械能，Ω 为角速度在 z 轴的投影，H 为动量矩在 ζ 轴的投影. 由这三个积分就可解得三个角 ψ, θ, φ 依赖于时间 t 的函数，需用椭圆积分表达.

考虑特殊情况，若 $\cos\theta=1$ 或 -1，则 $\theta=0$ 或 π，刚体的对称轴在铅直位置，前者不稳定，在干扰后会离开铅直位置使 $\theta\neq0$；后者总是稳定的. 若 $\cos\theta$ 为常数，则

$$\theta = \theta_0(常数),\ \frac{\mathrm{d}\psi}{\mathrm{d}t},\ \frac{\mathrm{d}\varphi}{\mathrm{d}t}$$

亦必为常数，刚体为规则进动.

一般情况下，若就其运动特征作定性分析，可令 $s=\cos\theta$，由以上第二、第三式解得

$$\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = c_1(1-s^2) - c_2 s(1-s^2) - c_3(c_4-s)^2 = f(s),$$
$$c_1 = \frac{2E-J_3\Omega^2}{J_1}, \quad c_2 = \frac{2mga}{J_1},$$
$$c_3 = \frac{J_3^2\Omega^2}{J_1^2}, \quad c_4 = \frac{H}{J_3\Omega}.$$

结合代数学中关于多项式根的理论以及物理上能够实现的情况，注意到 $s\to+\infty$ 使 $f(s)\to+\infty$，$s\to-\infty$ 使 $f(s)\to-\infty$，$f(1)<0$，$f(-1)<0$，以及 $|s|$

<1，需有 $f(s)\geqslant 0$ 才能使上列方程有意义. 故可推断三次多项式 $f(s)$ 的一个实根大于 1，另外的两个根应在 $(-1,1)$ 内，不妨设这二根 $s_1<s_2$，则物理上可实现的 s 必在 s_1 和 s_2 之间变化，记 $s_1=\cos\theta_1$，$s_2=\cos\theta_2$，则运动中章动角 θ 在 θ_2 和 θ_1 之间变化. 设 θ 不大于 $\pi/2$，则刚体的动力对称轴在两个圆锥面之间运动，该二圆锥面的顶角分别为 $2\theta_2$，$2\theta_1$. 若取以 O 为中心的球面截该圆锥面成二平行纬圈，则刚体运动中的重力对称轴就在此二纬圈间画出迹线. 为方便讨论，可认为 H，Ω 均为正值，由首次积分第二式得

$$\frac{\mathrm{d}\psi}{\mathrm{d}t}=\frac{1}{J_1\sin^2\theta}[H-J_3\Omega\cos\theta].$$

可见，当初始条件使 $\mathrm{d}\psi/\mathrm{d}t$ 不为 0，例如 $H/J_3\Omega>s_2$，则迹线如图 1；当初始条件使 $\mathrm{d}\psi/\mathrm{d}t$ 在 θ 处于 θ_1 和 θ_2 之间且等于 0，例如 $H/J_1\Omega$ 在 (s_1,s_2) 内，则迹线如图 2；当初始条件使 $\mathrm{d}\psi/\mathrm{d}t$ 在 $\theta=\theta_2$ 时且为 0，其他为正，迹线如图 3. 工程技术中的陀螺仪就是轴对称的匀质刚体做成的对称回转器，其自转角速度非常大. 忽略章动角速度和进动角速度，而把自转的动量矩近似地作为陀螺的总动量矩，形成了陀螺仪的近似理论，实用上有足够的精确度，可满足其应用.

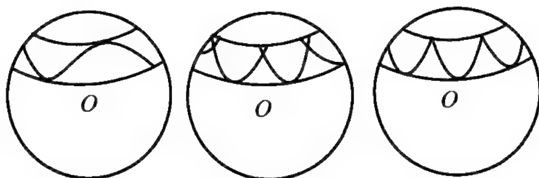


图 1　　图 2　　图 3

拉格朗日-泊松情形(Lagrange-Poisson condition)　即“拉格朗日情形”.

柯瓦列夫斯卡娅情形(Kovalevskaya condition)　任意初条件下，刚体绕定点运动可获解析解的一种情形. 即重刚体(所受主动力只是重力的刚体)对所绕固定点的惯量椭球(潘索椭球)为旋转对称的，刚体的质心位于赤道平面的情形. 若令对称轴为第三惯量主轴，另二惯量主轴在赤道平面内，则相应的转动惯量 $J_1=J_2=2J_3$. 俄国数学家柯瓦列夫斯卡娅(Ковалевская, C. B.)于 1889 年获此情形之解，故得名. 记固定点为 O，取 $O\xi\eta\zeta$ 为固连于地球的直角坐标系，$O\zeta$ 轴铅直向上；$Oxyz$ 为固连于运动刚体的直角坐标系，Oz 轴沿椭球对称轴，Ox 轴通过质心，a 为 O 到质心的距离. 欧拉动力学方程可写为

$$2\frac{\mathrm{d}\omega_1}{\mathrm{d}t}-\omega_2\omega_3=0,$$
$$2\frac{\mathrm{d}\omega_2}{\mathrm{d}t}+\omega_3\omega_1=\frac{Pa}{J_3}\gamma_3,$$
$$\frac{\mathrm{d}\omega_3}{\mathrm{d}t}=\frac{-Pa}{J_3}\gamma_2;$$

潘索方程为

$$\frac{\mathrm{d}\gamma_1}{\mathrm{d}t}=\omega_3\gamma_2-\omega_2\gamma_3,$$
$$\frac{\mathrm{d}\gamma_2}{\mathrm{d}t}=\omega_1\gamma_3-\omega_3\gamma_1,$$
$$\frac{\mathrm{d}\gamma_3}{\mathrm{d}t}=\omega_2\gamma_1-\omega_1\gamma_2.$$

式中 $\omega_1,\omega_2,\omega_3$ 依次为刚体角速度向量 $\boldsymbol{\omega}$ 在 x 轴、y 轴、z 轴的投影，$\gamma_1,\gamma_2,\gamma_3$ 依次为 $O\zeta$ 轴在 x 轴、y 轴、z 轴的方向余弦，P 为刚体重量. 除了机械能守恒：

$$J_3(2\omega_1^2+2\omega_2^2+\omega_3^2)+2Pa\gamma_1=\text{常量},$$

对 ζ 轴动量矩守恒：

$$2J_3\omega_1\gamma_1+2J_3\omega_2\gamma_2+J_3\omega_3\gamma_3=\text{常量},$$

以及 ζ 轴上单位向量长度的平方和 $\gamma_1^2+\gamma_2^2+\gamma_3^2=1$ 这样三个首次积分外，柯瓦列夫斯卡娅又求得第四个积分：将欧拉动力学方程的第二式乘以虚数 $\mathrm{i}=\sqrt{-1}$，并和第一个式子相加，得

$$2\frac{\mathrm{d}}{\mathrm{d}t}(\omega_1+\mathrm{i}\omega_2)=-\mathrm{i}\omega_3(\omega_1+\mathrm{i}\omega_2)+\mathrm{i}\frac{Pa}{J_3}\gamma_3.$$

也可将潘索方程作同样处理，得

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma_1+\mathrm{i}\gamma_2)=-\mathrm{i}\omega_3(\gamma_1+\mathrm{i}\gamma_2)+\mathrm{i}\gamma_3(\omega_1+\mathrm{i}\omega_2).$$

将所得二式分别乘以 $(\omega_1+\mathrm{i}\omega_2)$ 和 Pa/J_3，再相减以消去 γ_3，并整理为

$$\frac{\mathrm{d}}{\mathrm{d}t}\ln\left[(\omega_1+\mathrm{i}\omega_2)^2-\frac{Pa}{J_3}(\gamma_1+\mathrm{i}\gamma_2)\right]=-\mathrm{i}\omega_3.$$

用 $-\mathrm{i}=-\sqrt{-1}$ 代替 i，进行同样方法、步骤的运算，又得到

$$\frac{\mathrm{d}}{\mathrm{d}t}\ln\left[(\omega_1-\mathrm{i}\omega_2)^2-\frac{Pa}{J_3}(\gamma_1-\mathrm{i}\gamma_2)\right]=\mathrm{i}\omega_3.$$

将最后两式相加，经整理后有

$$\left(\omega_1^2-\omega_2^2-\frac{Pa}{J_3}\gamma_1\right)^2+\left(2\omega_1\omega_2-\frac{Pa}{J_3}\gamma_2\right)^2=\text{常数},$$

成为重刚体绕定点运动针对任意初条件可找到四个首次积分的三种情形之一. 最后可通过超椭圆积分表示该问题之解.

振动(vibration)　力学的基本概念之一. 指表征物质在某种状态附近往复变化的过程. 例如，物体位移在一定位置附近的往复变化称为机械振动；空间某点的电场强度和磁场强度随时间周期性的变化称为电磁振动(电磁振动常称为振荡). 一切波动，如声波、无线电波、光波等，都是某种振动的传播过程. 自然科学各学科存在各种各样的振动，本质上虽不相同，基本规律的描述却有许多共同之处，可用统一的方法处理. 一般说振动，常指机械振动，该系统总具有质量(惯性)和恢复力(弹性)，加上某种作用就形成振动. 振动可能发生于稳定平衡位置的附近，如单摆；也可能在定常运动的附近，如转轴的涡动. 工

程上前一情况居多，但两情况常可用同一方法处理，故常称振动为系统在平衡位置附近的往复运动. 按物理量随时间变化的规律来分，振动可分为周期振动（如简谐振动）和非周期振动（如衰减振动）. 按力学模型的自由度来分，可分为一个自由度系统的振动、多个自由度系统的振动和无限多自由度系统的振动. 前两种是具有集中参数的系统，对应着自变量为时间的常微分方程，最后一种对应着空间和时间为自变量的偏微分方程. 如一维的弦、杆振动，二维的薄膜振动，三维的弹性体一般振动. 按描述其规律的微分方程来分，可分为线性振动（方程是线性的）和非线性振动（方程是非线性的）. 其实，前者为后者的近似，后者是不宜于线性化的情形，其中还可能出现依赖于初条件而貌似随机的运动，称其为混沌. 若方程中不显含时间，则称该振动系统为自治的，方程中显含时间的为非自治的. 按振动起因来分，可分为自由振动、强迫振动、自激振动、参数振动. 第一种是仅仅受初始外来干扰而发生的振动，第二种是受随时间而变且独立于振动的外力作用，第三种是受到振动本身产生或控制的作用，第四种是系统的参数之一为时间的周期函数、频率适当时引起的系统振动. 按确定和不确定来分，又可分为确定性振动和随机振动. 前者可用确定性函数描述其规律，后者则需用概率统计方法进行定量描述，例如，风力引起的楼房振动就是这样. 研究振动，一般是要分析输入的激励、输出的响应以及振动系统本身的物理参量之间的关系，并用以指导实践.

电磁振动（electromagnetic vibration）　见“振动”.

振荡（oscillation）　即“电磁振动”.

自由振动（free vibration）　物体系不受外力的振动. 系统只受初始干扰发生的振动. 一般情况下，这种振动必因周围介质或系统本身的阻尼，使振幅不断减小呈衰减状态，称为有阻尼自由振动或衰减振动. 理想情况是过程中没有阻尼、能量守恒、振幅不变，称为无阻尼自由振动. 实际上，当阻尼很小、能量耗损不明显，在一定时间间隔内就可认为是无阻尼自由振动. 通常又把自由振动作为无阻尼自由振动的简称.

简谐振动（simple harmonic vibration）　亦称谐和振动或谐振动. 最简单、最基本的振动. 是可用时间的正（余）弦函数表达其振动变量的一类周期运动. 一切周期运动都可表示成这类振动的有限或无限的线性组合. 若不计阻力，单摆、复摆、扭摆、弹簧振子的微小振动都是简谐振动. 其运动微分方程为

$$\frac{d^2q}{dt^2} + \omega^2 q = 0,$$

q 是振动变量，对上述各摆动 q 为角度，对弹簧振子 q 为位移；t 为时间；ω 为角频率（圆频率），由振动系统本身的性质决定，故称为系统的固有角（圆）频率，有时也直接称为固有频率，是单位时间内振动位相的增量，单位是弧度/秒（rad/s）. 振动规律为 $q = A\sin(\omega t+\alpha)$（满足微分方程），称 A 为振幅，α 为初位相，α 由起始条件确定，$\omega t+\alpha$ 为位相. 在这种运动中，数值恒定的机械能和振幅的平方以及角频率的平方成正比. 当 q 为零时，广义速度最大，广义加速度为零；当 q 最大时，广义速度为零，广义加速度最大. 简谐振动是一个自由度、无阻尼、微小自由振动的理想化运动形态，通常把单摆的小振动作为这种运动的近似. 相应于数学上的傅里叶展开理论，对于任意的周期性振动，都可以表示为多种简谐振动的线性组合，因此，这种理想化模型有助于复杂振动形态的研究.

频率（frequency）　振动学概念. 简谐振动中单位时间内振动的次数. 单位是赫兹（Hz）（简称赫），等于次数或秒$^{-1}$，1kHz＝10^3Hz. 频率 f 和角频率 ω 的关系为 $2\pi f = \omega$.

圆频率（circular frequency）　见“简谐振动”.

角频率（angular frequency）　见“简谐振动”.

周期（period）　振动学概念. 周期振动中作一次完全振动所需的时间. 常用单位为秒（s）. 同一振动，其周期 T 和频率 f 的乘积等于 1，即 $Tf=1$.

固有频率（natural frequency）　一种特殊的频率. 指无阻尼线性系统固有振动的频率. 其个数与系统的自由度相同. 由小到大称为第一频率（一阶频率）、第二频率（二阶频率）、…，第一频率亦称为基本频率，简称基频. 一个自由度、无阻尼、线性的固有振动就是简谐振动，其固有频率即其频率. 由于角频率在运算上比较方便，使用较多，故亦直接称固有角频率为固有频率.

基本频率（basic frequency）　见“固有频率”.

主振动（principal vibration）　亦称固有振动. 振动学概念. 即无阻尼多自由度线性振动系统的单频率振动. 各广义坐标以同相或反相进行同频率的简谐振动. 这种频率的个数等于该系统的自由度，称为系统的主频率或固有频率. 故主振动的个数也等于系统的自由度. 只有在特殊条件下，系统才能实现其主振动. 一般初始条件下系统的自由振动是其各个主振动的线性组合.

固有振动（natural vibration）　即“主振动”.

阻尼（damping）　振动学概念. 振动系统所受的阻力. 该力来自系统内部、周围的介质或接触的物体，是实际振动无法避免的，将使系统的能量不断减少. 一般情况的阻尼与振动物体的速度成非线性关系，称该阻尼为非线性阻尼；在粘滞介质中小振动的阻尼可认为与速度成正比，称这种阻尼为粘滞阻尼

或线性阻尼. 合理地利用阻尼可减轻共振，保持系统的安全稳定.

非线性阻尼(non-linear damping) 见"阻尼".

线性阻尼(linear damping) 见"阻尼".

阻尼振动(damped vibration) 有阻尼的振动. 由于阻尼的影响使能量不断减少的自由振动. 一切实际的振动总受有系统本身及周期介质的阻力(在振动中称为阻尼)，振动系统需克服阻力作功，且系统对周围弹性介质的作用将以波动形式传播，这些就使能量不断减少，振幅呈现衰减，故亦称衰减振动. 例如，弹簧-质量-阻尼系统，设 k 为弹簧系数，m 为质量，μ 为阻力系数，则其运动微分方程为

$$\frac{d^2x}{dt^2}+2n\frac{dx}{dt}+\omega_0^2x=0,$$

式中 $\omega_0^2=k/m$，ω_0 为系统的固有角频率，$n=\mu/2m$ 为阻尼系数. 当阻尼小时，即 $n<\omega_0$ 时，称为欠阻尼，则

$$x=Ae^{-nt}\sin(\sqrt{\omega_0^2-n^2}t+\alpha),$$

A 和 α 为由初条件和系统特性决定的常量. 这就是一种幅值 Ae^{-nt}按指数规律减小的衰减振动. 它不是周期运动，有时称为伪简谐运动，习惯上仍称相邻两峰值的时间间隔 $T=2\pi/\omega$ 为周期，其中

$$\omega=\sqrt{\omega_0^2-n^2}=\omega_0\sqrt{1-\zeta^2}$$

为角频率，称 $\zeta=n/\omega_0$ 为阻尼比. 可见，阻尼使自由振动的频率减小，周期增大. 相隔时间为 T 的两相邻振幅之比为 e^{nT}，可表明振幅的衰减情况，称

$$nT=\frac{2\pi n}{\sqrt{\omega_0^2-n^2}}=\frac{2\pi\zeta}{\sqrt{1-\zeta^2}}$$

为对数减缩或对数减幅系数. 实用中可通过振幅比和 T 的测量求 n. 若阻尼较大，使 $n>\omega_0$，称为过阻尼，则运动不是振动，位移最多有一次达到极大值，需长时间才能慢慢回到平衡位置，n 越大越慢. 若 $n=\omega_0$ 称为临界阻尼，运动也不是振动，而使系统刚刚能不作往复运动，而最快地回到平衡位置. 实用中，如仪表中的阻尼器就是采用临界阻尼.

阻尼比(damping ratio) 见"阻尼振动".

阻尼系数(damping coefficient) 见"阻尼振动".

临界阻尼(critical damping) 阻尼的一种特殊情况. 是使不受强迫力作用的系统受初干扰后，可不经往复运动回到平衡位置的最小粘滞阻尼(参见"阻尼振动").

衰减振动(attenuation vibration) 即"阻尼振动".

欠阻尼(underdamping) 见"阻尼振动".

过阻尼(overdamping) 见"阻尼振动".

受迫振动(forced vibration) 亦称强迫振动. 系统受外力干扰下的振动. 振动系统对所受激励的响应. 例如，机器运转时偏心转子引起整个机器的振动. 随时间而变化的强迫力(干扰力)是常见的激励，可有非周期性和周期性两大类，后者又包括一般周期性和简谐周期性两种. 以质量-弹簧-阻尼器为典型的单自由度系统为例，令 x,t 分别表示位移、时间，对依赖于时间的强迫力$f(t)$，其运动微分方程为

$$\frac{d^2x}{dt^2}+2n\frac{dx}{dt}+\omega_0^2x=\frac{f(x)}{m},$$

式中 $n=\mu/2m,\omega_0^2=k/m,\mu$ 为阻力系数，m 为质量，k 为弹簧系数，ω_0 为系统的固有角频率. 在欠阻尼($n<\omega_0$)下，有

$$x=e^{-nt}\Big[x_0\cos\sqrt{\omega_0^2-n^2}\,t+\frac{nx_0+\dot{x}_0}{\sqrt{\omega_0^2-n^2}}\cdot\sin\sqrt{\omega_0^2-n^2}\,t\Big]+\frac{1}{m\sqrt{\omega_0^2-n^2}}\cdot\int_0^t\Big[f(s)e^{-n(t-s)}\sin\sqrt{\omega_0^2-n^2}(t-s)ds\Big].$$

若 $f(t)=H\sin pt$ 为简谐的周期函数，则

$$x=Ae^{-nt}\sin(\sqrt{\omega_0^2-n^2}\,t+\alpha)+B\sin(pt-\beta),$$

A,α 由初条件确定，

$$B=\frac{H}{m}\frac{1}{\sqrt{(\omega_0^2-p^2)^2+4n^2p^2}},$$

$$\beta=\arctan\frac{2np}{\omega_0^2-p^2}.$$

解的第一部分为阻尼振动，必随时间的增长而逐渐消失；第二部分是与强迫力同周期的简谐振动，称为稳定受迫振动或纯受迫振动. 其振幅、初位相不依赖于初条件，而取决于强迫力和振动系统本身. 可将 B 表示为

$$B=\frac{B_0}{\sqrt{(1-z^2)^2+4\nu^2z^2}},$$

$B_0=H/k$ 为静力偏移，$z=p/\omega_0$，$\nu=n/\omega_0$，称

$$\lambda=\frac{1}{\sqrt{(1-z^2)^2+4\nu^2z^2}}$$

为动力放大系数. 工程技术中很关注 λ 随 z 变化的情况，称 $\lambda(z)$(ν 为其参数)为幅频特性曲线或共振曲线. 对于 $f(x)$是任意周期性力的情形，可展开为傅里叶级数，根据振动系统的线性性质，利用叠加原理可求其解，只需 $f(x)$ 满足一般较易具备的狄利克雷条件.

纯受迫振动(pure forced vibration) 见"受迫振动".

共振(resonance) 一种振动现象. 指受迫振动的振幅达到最大值的现象. 当振动系统所受激励的频率接近其固有频率时，即发生这种现象. 对质量-

弹簧-阻尼器这种典型的振动系统，若受到简谐强迫振动的频率为 p，在

$$p=\sqrt{\omega_0^2-2n^2}$$

时，就使纯受迫振动的振幅 B 达到最大值，即

$$B_{最大}=\frac{H}{2mn\sqrt{\omega_0^2-n^2}},$$

ω_0 为系统的固有角频率，m 为质量，H 为强迫力的最大值，$n=\mu/2m$，μ 为阻力系数，n 为阻尼系数. 当阻尼很小时，引发共振的强迫力频率即接近固有频率 ω_0，共振的振幅接近 $H/2mn\omega_0$. 当 n 趋近于零时，共振的振幅趋于无穷，当然，n 不可能等于零，否则纯受迫振动也不会出现. 若上述振动不是小振动，则所讨论系统已不能体现实际的物理性质，需要非线性理论. 共振是实践中常见且重要的现象，机械的、声的、磁电的各领域都存在共振现象. 有的情况需要利用共振，如乐器音响的提高、收音机的调谐、共振筛的应用；有时又需要避免共振，如建筑结构和机器设备就力避共振，以保证稳定和安全.

摆(pendulum)　一类装置. 是能产生往复摆动的一类装置. 将无重细杆或不可伸长的细柔绳一端悬于重力场内一定点，另一端固结一个重小球，就构成数学摆，亦称单摆. 若小球只限于铅直平面内摆动，称为平面数学摆. 若小球摆动不限于铅直平面，则称为球面数学摆. 在重力场内装置一个水平轴，使具有一定质量的刚体能绕该轴光滑地转动，则形成一个物理摆，亦称复摆. 若计入地球自转的影响，足够长的单摆小摆动又称为佛科摆. 重力场内悬于空间的金属细轴下端固连一个刚体，在受到绕轴转动的干扰后，能发生往复扭动，就构成一个扭摆. 各种摆动都属于机械振动，一般情况都是非线性的，只当振动足够小时，才能线性化且呈简谐形式. 过悬挂点 O 取直角坐标系 $Oxyz$，令 Oz 轴铅垂向下，为表明佛科摆设在北半球，令 Ox 轴向东，Oy 轴向南，ψ 为纬度，ω 为地球角速度，平面数学摆的运动方程为

$$m\frac{\mathrm{d}v}{\mathrm{d}t}=-mg\sin\varphi,$$

$$m\frac{v^2}{l}=-mg\cos\varphi+N;$$

球面数学摆的运动方程为

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=-N\frac{x}{l},$$

$$m\frac{\mathrm{d}^2y}{\mathrm{d}t^2}=-N\frac{y}{l},$$

$$m\frac{\mathrm{d}^2z}{\mathrm{d}t^2}=mg-N\frac{z}{l};$$

物理摆的运动方程为

$$J\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2}=-mgl\sin\varphi;$$

佛科摆的运动方程为

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=-N\frac{x}{l}-2m\omega\left(-\cos\psi\frac{\mathrm{d}z}{\mathrm{d}t}+\sin\psi\frac{\mathrm{d}y}{\mathrm{d}t}\right),$$

$$m\frac{\mathrm{d}^2y}{\mathrm{d}t^2}=-N\frac{y}{l}+2m\omega\sin\psi\frac{\mathrm{d}x}{\mathrm{d}t},$$

$$m\frac{\mathrm{d}^2z}{\mathrm{d}t^2}=mg-N\frac{z}{l}-2m\omega\cos\psi\frac{\mathrm{d}x}{\mathrm{d}t};$$

扭摆的运动方程为

$$I\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}+C\theta=0.$$

各式中 m 为质量，g 为重力加速度，v 为速度，t 为时间，l 为质心到定点(轴)的距离，φ 为摆动偏角，θ 为扭转角，N 为张力，C 为抗扭刚度，J 为对转轴的转动惯量，I 为对扭轴的转动惯量. 对于小振动情况下的运动规律，平面数学摆为

$$\varphi=A\sin\left[\sqrt{\frac{g}{l}}\,t+\alpha\right],$$

α，A 由初条件决定，初位移和初速度共线；球面数学摆为

$$x=B\sin\left[\sqrt{\frac{g}{l}}t+\beta\right],\ y=C\sin\left[\sqrt{\frac{g}{l}}\,t+\gamma\right],$$

β，γ，B，C 由初条件决定，初位移和初速度不共线；物理摆为

$$\varphi=D\sin\left(\sqrt{\frac{mgl}{J}}t+\delta\right),$$

D，δ 由初条件决定；佛科摆摆面为

$$\theta=\theta_0+(\omega\sin\psi)t,$$

θ 是摆动平面与 Ox 的夹角，θ_0 为初始值，可得摆动平面的转动角速度为 $\omega\sin\psi$，在北半球的转向是由东到南再到西，转动一周所需时间为 $T=2\pi/\omega\sin\psi$ 或 $24/\sin\psi$ 小时；扭摆为

$$\theta=E\sin\left[\sqrt{\frac{C}{I}}t+\varepsilon\right],$$

E 和 ε 由初条件决定.

摆的运动在历史上曾引起不少科学家的兴趣. 意大利科学家伽利略(Galilei，G.)曾研究过单摆小振动下的等周期运动. 从那时以后，摆被用来制造计时的钟表. 后来荷兰数学家、物理学家惠更斯(Huygens，C.)研究过钟摆的运动特性. 英国数学家、物理学家牛顿(Newton，I.)也研究过摆的运动规律. 佛科(Foucault，J. L.)于 1851 年在巴黎用 67m 长的单摆进行了出色的实验，验证了地球的自转，博得很高声誉. 这种计入地球自转的大型单摆就称为佛科摆. 后人多次重复这种实验，并加以改进. 世界上有不少天文馆都装有一长钢丝下悬钢球的佛科摆. 在北京地区，$\psi=40°$，摆面转动周期为 $T=37.34$h，这种因地球自转带来的摆平面以角速度 $\omega\sin\psi$ 转动的效应，称为佛科效应.

单摆(simple pendulum) 见"摆".

数学摆(mathematical pendulum) 见"摆".

复摆(compound pendulum) 见"摆".

物理摆(physical pendulum) 见"摆".

扭摆(torsional pendulum) 见"摆".

佛科摆(Foucault pendulum) 见"摆".

佛科效应(Foucault effect) 见"摆".

碰撞(impact) 一种特殊的动力学现象.指在极短时间(约千分之几到万分之几秒)内,经物体之间接触,使机械运动特征量突然改变的现象,如打桩、冲压加工、踢球、交通事故中汽车相撞等.其特点为:

1.碰撞中变形效应不可忽略,碰撞物体适合弹塑性体模型.但情况复杂,理论力学中不考虑变形的具体过程,只计入它对机械运动影响的数值.

2.碰撞力(亦称冲击力,瞬时力)远远大于一般的非碰撞力(数百或数千倍),在研究中忽略非碰撞力.

3.碰撞前后物体位置的改变很小,可忽略不计.

4.过程中碰撞力变化急剧.

因此,考虑力对碰撞物体效应时,使用其冲量、冲量矩,并运用积分形式的动量定理和动量矩定理来分析问题.例如,质点系的动量在碰撞中的改变量,等于所受外碰撞冲量的主向量,即

$$\Delta\sum m\boldsymbol{v} = \Delta(M\boldsymbol{v}_c) = \sum \boldsymbol{s},\ \sum m\boldsymbol{v} = M\boldsymbol{v}_c,$$

$M\boldsymbol{v}_c$ 为质点系的动量,M 为总质量,$\boldsymbol{v}_c$ 为质心速度,$\boldsymbol{s}$ 为外碰撞冲量;质点系对某定点的动量矩在碰撞中的改变量,等于所受外碰撞冲量对该点的主矩,即

$$\Delta\sum(\boldsymbol{r}\times m\boldsymbol{v}) = \sum(\boldsymbol{r}\times\boldsymbol{s}),$$

$\boldsymbol{r}$ 为质点系中质点相对定点的向径;质点系相对于随质心平动坐标的运动,对质心的动量矩在碰撞中的改变量,等于所受外碰撞冲量对质心的主矩,即

$$\Delta\sum(\boldsymbol{r}'\times m\boldsymbol{v}') = \Delta\sum(\boldsymbol{r}'\times\boldsymbol{s}),$$

$\boldsymbol{r}'$,$\boldsymbol{v}'$ 分别为质点系中质点相对质心的向径、速度.若质点系中只有一个质点,则动量定理表示为

$$\Delta(m\boldsymbol{v})=\boldsymbol{s},$$

动量矩定理表示为

$$\Delta(\boldsymbol{r}\times m\boldsymbol{v})=\boldsymbol{r}\times\boldsymbol{s}.$$

对各种运动形式的刚体碰撞,可用类似方法处理.在质点的碰撞中,用恢复系数来表征变形的影响,以弥补刚性模型的不足.在具体的物体碰撞中,这种系数有一定参考价值.

碰撞力(impact force) 见"碰撞".

瞬时力(instantaneous force) 即碰撞力.见"碰撞".

恢复系数(coefficient of restitution) 描述物体碰撞的参数.是表征碰撞物体变形恢复程度的参数.主要取决于材料的性质,也与碰撞物的其他因素如形状、大小、相对接触点公法向的速度等有关.恢复系数可由实验确定,将两种材料制成的小球沿连心线运动而相撞.设 $\boldsymbol{v}_1$,$\boldsymbol{v}_2$ 和 $\boldsymbol{u}_1$,$\boldsymbol{u}_2$ 分别为两小球在碰撞前和碰撞后的速度,则恢复系数

$$k = \left|\frac{\boldsymbol{u}_2-\boldsymbol{u}_1}{\boldsymbol{v}_2-\boldsymbol{v}_1}\right|,$$

可推得 $k=s_2/s_1$,其中 s_1,s_2 分别为碰撞中变形和恢复阶段的冲量.对于同平面内运动的两球斜碰撞,$\boldsymbol{v}_1$,$\boldsymbol{v}_2$,$\boldsymbol{u}_1$,$\boldsymbol{u}_2$ 应看做是两球碰撞前和碰撞后在接触点公法向的速度.若把两种材料分别制成小球和固定的平板,令小球自高度 h_1 落于平板上,回弹高度为 h_2,则

$$k = \sqrt{\frac{h_2}{h_1}},$$

这是最简单的一种测定恢复系数的方法.对于完全弹性材料,$k=1$,表明变形可全部恢复,相应的碰撞称为弹性碰撞.对于完全塑性材料,$k=0$,表明变形全部不恢复,相应的碰撞称为完全塑性碰撞或非弹性碰撞.一般材料的 k 在 0 与 1 之间.例如,木对木平均为 0.5,钢对钢为 0.56,玻璃对玻璃为 0.95,铁对铅为 0.14.这些材料的碰撞称为不完全的弹性碰撞,碰撞后变形只部分恢复.英国数学家、物理学家牛顿(Newton,I.)曾在实验基础上假定 k 值是材料性质决定的常数,也称该系数为材料的恢复系数,这与上面指出的近代结论有一定出入.所以以上针对小球测得的数据,对于具体物体只有参考价值.实用中,运用恢复系数估算碰撞前后机械运动量的传递、动能的损耗,十分简捷,应该合理地运用这种系数.

弹性碰撞(elastic impact) 见"恢复系数".

完全塑性碰撞(perfect plastic impact) 见"恢复系数".

不完全弹性碰撞(non-perfect elastic impact) 见"恢复系数".

撞击中心(center of percussion) 动力学术语.指某种定轴转动体上的一个点,若通过该点沿某一作用线给物体施冲量,无论大小如何,均使转轴不感受冲击力.通常,绕轴转动的物体受撞击后,将把冲击力传于转轴,若转轴是物体相对于轴上一点 O 的惯量主轴,则存在撞击中心 P.施于物体的冲量通过 P,且满足:

1.在过 O 而垂直于轴的平面 S_1 内;

2.垂直于质心 C 和转轴决定的平面 S_2;

则可使转轴免受冲力.即 P 在 S_1,S_2 的交线上,和 C 位于轴同侧,且

$$|OP| = \frac{J}{M\delta},$$

J 为物体对轴的转动惯量，M 为质量，δ 为 C 到轴的距离.

对可化为平面图形，且绕定点 O 在自身平面内转动的情形，例如有对称平面的匀质体绕垂直轴转动，则 P,C 在 O 的同侧，满足

$$|OP| \cdot |OC| = \rho^2,$$

ρ 为图形对 O 的惯量半径. 转轴免受冲力的条件是：施于图形的冲量通过 P，并在图形平面内垂直于 O，C 的连线. 实践中，如使用榔头撞击试验机的摆锤等碰撞器具，常可因力点和支点的位置符合撞击中心理论，使支点不受冲击力.

流体力学

流体(fluid)　物质的一种形态. 是静止时不能承受剪切应力作用的物质. 液体和气体是流体. 流体力学中有以下几种重要的流体模型：

1. 牛顿流体. 它是指本构关系为

$$\tau_{ij} = \mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3}\frac{\partial u_\alpha}{\partial x_\alpha}\delta_{ij}\right) + \mu' \frac{\partial u_\alpha}{\partial x_\alpha}\delta_{ij}$$

的流体，这里采用了约定求和记号，$x_i(i=1,2,3)$是直角坐标，τ_{ij}和 u_i 分别是粘性应力张量和速度在此坐标系下的分量，δ_{ij}是克罗内克记号，μ 和 μ' 分别是第一和第二粘性系数，它们是物性常数，只取决于流体所处的热力学状态(即为压力和温度的函数). 当流体分子没有内自由度或内自由度没有激发时，第二粘性系数 $\mu'=0$；在许多流体力学问题中，流体的体膨胀率(即速度的散度$\frac{\partial u_\alpha}{\partial x_\alpha}$)与变形速度张量

$$\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$$

中的一些分量相比是很小的，因而第二粘性系数的作用也是很小的. 对这类流体力学问题，也可认为第二粘性系数 $\mu'=0$. 第二粘性系数等于零的牛顿流体也称为斯托克斯流体.

2. 不可压缩流体. 指任一部分体积在运动中恒保持不变的流体. 在数学上可表示为：速度的散度$\frac{\partial u_\alpha}{\partial x_\alpha}$处处都等于零.

3. 正压流体. 指密度 ρ 可以由压力 p 惟一决定的流体. 其状态方程可写成 $\rho=\rho(p)$.

4. 完全气体. 指状态方程可写成 $p/\rho=RT$ 的流体，这里 T 是绝对温度，R 是气体常数. 完全气体的内能、焓、定容比热和定压比热都可以是温度的函数，但许多情况下，可把气体看做是比热为常数的完全气体.

5. 理想流体. 指完全忽略粘性，即粘性系数为零的流体.

6. 非牛顿流体. 指粘性应力张量与变形速度张量间不成正比的流体.

流体静力学(hydrostatics)　流体力学的一个分支学科. 流体静力学研究外力作用下流体相对静止时的平衡条件、有关物理量在流体中的分布及流体与边界之间的相互作用. 研究流体静力学问题可以以刚化原理为基础. 此原理认为：把流体中的任一部分刚化后，不会破坏流体原先的相对静止状态. 相对静止的流体中，任一截面上的剪切应力都等于零，因此流体中的应力张量一定是球对称的. 各分量可写成

$$\sigma_{ij} = -p\delta_{ij},$$

δ_{ij}是克罗内克记号，p 是压力(流体力学中，压力与压强的含义是相同的).

平衡方程(equilibrium equation)　流体的平衡条件. 表示作用在任一部分静止流体上的，包括来自周围流体对它的作用在内的所有外力应平衡的关系式. 平衡方程的微分形式是$\rho\boldsymbol{f}=\mathrm{grad}\,p$，$\boldsymbol{f}$ 是作用在单位质量流体上的外力，称为质量力，ρ 和 p 分别是流体的密度和压力. 从平衡方程可得出，质量力应满足 $\boldsymbol{f}\cdot\mathrm{rot}\,\boldsymbol{f}=0$，这是一个相容方程，如果外力不满足这一条件，流体将不可能相对静止. 相容方程的特殊情况是 $\mathrm{rot}\,\boldsymbol{f}=0$，这时，质量力是有势的，$\boldsymbol{f}=-\mathrm{grad}U$，$U$ 是质量力的势函数. 有势的质量力称为保守力.

正压流体的平衡(equilibrium of barotropic fluid)　正压流体的平衡条件. 这里是指正压流体(参见“流体”)相对于某参考系为静止的状态. 正压流体平衡时，作用在流体上的质量力 $\boldsymbol{f}$ 一定是有势的，$\boldsymbol{f}=-\mathrm{grad}U$，$U$ 是势函数. 流体中压力 p 的分布为

$$\int\frac{\mathrm{d}p}{\rho} + U = \text{常数},$$

其中$\int\frac{\mathrm{d}p}{\rho}$是压力函数，$\rho$ 是流体密度. 在密度均匀的情况下，有

$$\frac{p}{\rho} + U = \text{常数}.$$

在重力场中，质量力的一般形式是 $\boldsymbol{f}=\boldsymbol{g}-\boldsymbol{w}$，$\boldsymbol{g}$ 是重力加速度，$\boldsymbol{w}$ 是牵连加速度. 为使 $\boldsymbol{f}$ 有势，并且与时间无关，参考系必定是作等加速直线运动的，或是绕定轴作等速转动的，或是这两种运动形式的组合. 这时，$\boldsymbol{f}$ 的势函数 U 可写成

$$U = (\boldsymbol{a}_0 - \boldsymbol{g})\cdot\boldsymbol{r} - \frac{\omega^2}{2}R^2,$$

$\boldsymbol{a}_0$ 是参考系原点的平动加速度，$\boldsymbol{r}$ 是空间点对原点的向径，ω 是参考系的角速度，R 是空间点到旋转轴的距离.

帕斯卡定律(Pascal law)　流体静力学中的一

条基本定律.帕斯卡定律指出:不可压缩静止流体中,任一处因某种外因而产生的压力增值,会立即传到流体各点.帕斯卡定理表明:在不可压缩流体中,压力扰动传播的速度为无穷大.帕斯卡定律由法国数学家帕斯卡(Pascal,B.)于1653年首先提出而得名.

阿基米德定律(Archimedes law) 流体静力学的一条基本定律.是关于浸没于静止流体或浮于静止流体表面的物体受到流体合作用力(浮力)的一条定律.阿基米德定律指出:静止流体对浸没于流体或浮于流体表面的物体的合作用力,与被物体排开的那部分流体的质量力的合力大小相等、方向相反.利用刚化原理(参见“流体静力学”)可直接得到阿基米德定律.阿基米德定律由古希腊科学家阿基米德(Archimedes)提出.

流体运动学(fluid kinematics) 流体力学的一个分支学科.它研究流体运动的几何性质,不考虑力的具体作用和流体运动的原因.流体运动学的基本内容有:流场的几何表示,流体物理量的随体导数,流体微团的速度分解,涡量场的运动学性质以及涡量场与速度场之间的相互联系等.

流体运动的描述方法(description method of fluid motion) 描述流体运动的两种方法,即拉格朗日描述方法与欧拉描述方法.拉格朗日方法描述每个流体质点的位置随时间的变化.欧拉方法描述每一时刻空间各点上流体质点的速度.除运动学的量以外,对压力、密度、温度等物理量也有类似的两种描述方法.拉格朗日方法描述的是随体量,欧拉方法描述的是有关物理量的场.根据质量、动量和能量的守恒定律,可导出两种描述方法的流体力学基本方程组.

平面流动(plane flow) 一类特殊流动.如果各点的流速都平行于某一平面,并且所有物理量在此平面的垂直方向上是不变的,则称此流动是平面流动.对平面流动,总可以找到某直角坐标系 $Oxyz$,在此坐标系中,沿 z 轴的速度分量为零,并且所有物理量对 z 的偏导数均为零.

轴对称流动(axisymmetric flow) 一类特殊的流动.如果能找到一个圆柱坐标系 $O\rho z\theta$,在此坐标系中,沿 θ 方向的速度分量恒等于零,并且所有物理量对 θ 的导数均为零,则称此流动为轴对称流动.在轴对称流动中,任一过轴线 $z=0$ 的平面(子午面)中的流体质点将永远在此平面内,并且所有的物理量沿此平面法线方向的导数均为零.

随体导数(material derivative) 一种变化率.它是从属于具体流体的物理量在流动过程中随时间的变化率.物理量 $\boldsymbol{F}$ 的随体导数为

$$\frac{\mathrm{D}\boldsymbol{F}}{\mathrm{D}t}=\frac{\partial \boldsymbol{F}}{\partial t}+(\boldsymbol{v}\cdot\nabla)\boldsymbol{F},$$

$\boldsymbol{v}$ 为速度,t 为时间,$\frac{\partial \boldsymbol{F}}{\partial t}$ 称为 $\boldsymbol{F}$ 的当地导数,$(\boldsymbol{v}\cdot\nabla)\boldsymbol{F}$ 称为 $\boldsymbol{F}$ 的对流导数.速度 $\boldsymbol{v}$ 的随体导数,即加速度为

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}=\frac{\partial \boldsymbol{v}}{\partial t}+(\boldsymbol{v}\cdot\nabla)\boldsymbol{v}.$$

随体导数与通常的偏导数一般不能交换次序.例如

$$\frac{\partial}{\partial t}\left(\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}\right),\quad \nabla\cdot\left(\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}\right)\text{和}\nabla\times\left(\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}\right)$$

并不等于

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\partial \boldsymbol{v}}{\partial t}\right),\ \frac{\mathrm{D}}{\mathrm{D}t}(\nabla\cdot\boldsymbol{v})\text{和}\frac{\mathrm{D}}{\mathrm{D}t}(\nabla\times\boldsymbol{v}),$$

而有

$$\frac{\partial}{\partial t}\left(\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}\right)=\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\partial \boldsymbol{v}}{\partial t}\right)+\frac{\partial u_\alpha}{\partial t}\frac{\partial \boldsymbol{v}}{\partial x_\alpha},$$

$$\nabla\cdot\left(\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}\right)=\frac{\mathrm{D}}{\mathrm{D}t}(\nabla\cdot\boldsymbol{v})+\frac{\partial u_i}{\partial x_\alpha}\frac{\partial u_\alpha}{\partial x_i},$$

$$\nabla\times\left(\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}\right)=\frac{\mathrm{D}}{\mathrm{D}t}(\nabla\times\boldsymbol{v})-(\mathrm{rot}\boldsymbol{v}\cdot\nabla)\boldsymbol{v}+(\nabla\ \boldsymbol{v})+\mathrm{rot}\boldsymbol{v},$$

这里 u_i 是 $\boldsymbol{v}$ 的三个速度分量.某一部分流体体积 τ 中物理量 $\boldsymbol{F}$ 的总和,即体积分 $\int_\tau \boldsymbol{F}\mathrm{d}\tau$ 的随体导数为

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_\tau \boldsymbol{F}\mathrm{d}\tau=\int_\tau \frac{\partial \boldsymbol{F}}{\partial t}\mathrm{d}\tau+\oint_A \boldsymbol{F}v_n\mathrm{d}A,$$

式中 A 是体积 τ 的边界面,v_n 是 A 上流体质点在 A 的外法线方向上的速度分量.

流线(stream line) 一条曲线.指某时刻空间中的一条曲线,此曲线上各点的切线与该点处流体速度的方向重合.如果 $\boldsymbol{v}$ 为速度场,则确定流线的方程可写成 $\boldsymbol{v}\times\mathrm{d}\boldsymbol{l}=0$,$\mathrm{d}\boldsymbol{l}$ 是流线上的向量微元.流线与流体质点运动的历史无关,只决定于某瞬时空间各点的流速方向.流线与流体质点的轨迹(迹线)是不同的,后者决定于具体流体质点的运动历史.流线与从空间某点流出的染色流体质点在某时刻组成的曲线(烟线,也称为脉线)是不同的.通过某不自交的非流线的封闭曲线上各点的流线所组成的管状曲面称为流管.

流管(stream tube) 见“流线”.

速度环量(velocity circulation) 表征流动状态的一种物理量.速度 $\boldsymbol{v}$ 沿某一封闭曲线 l 的环量是指积分

$$\Gamma=\oint_l \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{l},$$

$\mathrm{d}\boldsymbol{l}$ 是曲线 l 上的向量微元.

速度环量可用来说明涡旋运动的某些性质.说明涡旋运动的另一个量是涡量,它是 $\Omega=\mathrm{rot}\boldsymbol{v}$.涡量场是向量场.空间某点的涡量等于该点处流体微团刚化后角速度的两倍.速度环量与涡量之间由斯托

克斯公式相联系

$$\oint_l \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{l} = \int_A \boldsymbol{n} \cdot \mathrm{rot}\boldsymbol{v}\mathrm{d}A,$$

A 是以封闭曲线 l 为周界的曲面，$\boldsymbol{n}$ 是曲线 A 上的单位法向量，其指向可根据线积分的走向由右螺旋的进动方向决定.

涡量(vorticity)　见“速度环量”.

涡线(vortex line)　流体静力学的基本概念之一. 指某时刻处处与涡量相切的曲线. 如果 $\mathrm{d}\boldsymbol{l}$ 是涡线上的微元向量，$\boldsymbol{\Omega}$ 是涡量，则确定涡线的微分方程是 $\boldsymbol{\Omega}\times\mathrm{d}\boldsymbol{l}=0$. 通过非涡线且不自交的封闭曲线各点的所有涡线所组成的管状曲面称为涡管. 涡管某截面 A 上的涡通量为

$$\int_A \boldsymbol{n} \cdot \boldsymbol{\Omega}\mathrm{d}A,$$

$\boldsymbol{n}$ 是截面 A 的单位法向量，其方向沿此截面上涡量所指的方向. 此涡通量称为涡管强度. 由于涡量 $\boldsymbol{\Omega}$ 的散度一定等于零，涡管强度沿涡管是不变的. 涡管不能在流体内部中断. 对真实流体来说，涡管也不能起迄于静止的固体边界.

涡管(vortex tube)　见“涡线”.

速度分解定理(velocity decomposition theorem)　关于流体微团速度分解的一条基本定理：如果某点 O 处的速度为 $\boldsymbol{v}_O$，则与 O 点邻近的 P 点处的速度 $\boldsymbol{v}$ 可表示成

$$\boldsymbol{v}=\boldsymbol{v}_O+\frac{1}{2}\mathrm{rot}\ \boldsymbol{v}\times\delta\boldsymbol{r}+\delta\boldsymbol{r}\cdot\boldsymbol{S},$$

式中 $\delta\boldsymbol{r}$ 是从 O 点到 P 点的向量，$\boldsymbol{S}$ 是速度导数张量的对称部分，其各分量为

$$s_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right),$$

$\boldsymbol{S}$ 称为变形速度张量. 对刚体来说，也有类似的速度分解，但变形速度张量恒为零，角速度 $\mathrm{rot}\boldsymbol{v}/2$ 在刚体各点是相同的. 流体的速度分解只能在流体中某点的小邻域内进行：此点小邻域内流体速度可分解为类似于刚体运动部分

$$\boldsymbol{v}_O+\frac{1}{2}\mathrm{rot}\boldsymbol{v}\times\delta\boldsymbol{r}$$

和变形部分 $\delta\boldsymbol{r}\cdot\boldsymbol{S}$.

无旋运动(irrotational flow)　一种特殊流动. 涡量处处等于零的流动称为无旋运动. 在无旋运动中，必存在某标量场 φ，其梯度等于流速，$\boldsymbol{v}=\mathrm{grad}\varphi$，$\varphi$ 称为速度势. 速度势 φ 也可写成

$$\varphi=\int_A^B \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{l},$$

这里 A 是一个定点，可规定该点处的速度势的值为零；B 是动点，积分沿任何连结 A 和 B 两点的曲线进行；$\mathrm{d}\boldsymbol{l}$ 是曲线上的微元向量.

速度势(velocity potential)　见“无旋运动”.

流函数(stream function)　与质量守恒紧密联系的函数. 在不可压缩流体的运动中，速度 $\boldsymbol{v}$ 需满足方程 $\nabla\cdot\boldsymbol{v}=0$. 因此必存在某向量函数 $\boldsymbol{B}$，使得 $\boldsymbol{v}=\nabla\times\boldsymbol{B}$，$\boldsymbol{B}$ 称为广义流函数. 在平面流动中，如取直角坐标系，有

$$\boldsymbol{B}=(0,0,\psi),\ u=\frac{\partial\psi}{\partial y},\ v=-\frac{\partial\psi}{\partial x},$$

u 和 v 是速度 $\boldsymbol{v}$ 在 x 方向和 y 方向的分量. 函数 ψ 称为拉格朗日流函数. 拉格朗日流函数有以下性质：

1. 曲线 $\psi(x,y)=$ 常数，是一条流线；

2. 流函数沿某一方向的导数，等于从此方向顺时针旋转 90°角方向上的速度分量；

3. 两不同点上流函数之差，等于连结此两点的曲线(单位厚度的柱面)上的体积流量.

在不可压缩流体的轴对称流动中，流函数 ψ 与广义流函数 $\boldsymbol{B}$ 之间，在圆柱坐标系下有

$$\boldsymbol{B}=\left(0,0,\frac{\psi}{r}\right),$$

r 是空间点到对称轴的距离. 而在球坐标系下，则有

$$\boldsymbol{B}=\left(0,0,\frac{\psi}{r\sin\theta}\right),$$

r 是空间点到原点的距离，θ 是纬度角. 这时的流函数 ψ 称为斯托克斯流函数. 在可压缩流体的定常流动中，广义流函数 $\boldsymbol{B}$ 与速度 $\boldsymbol{v}$ 之间的联系是 $\rho\boldsymbol{v}=\nabla\times\boldsymbol{B}$，$\rho$ 是流体密度.

不可压缩流体的无旋运动(irrotational flow of incompressible fluid)　一类具有重要实际背景的特殊流动. 在这类流动中，流体的压缩性可以忽略，并且涡量处处等于零，因而存在速度势 φ，满足拉普拉斯方程 $\nabla^2\varphi=0$. 根据调和函数的性质，速度势 φ 不能在流体内部取得极值，因此在下述的任一种条件下，均可惟一确定有界单连通区域中的速度势：

1. 在边界上给定速度势 φ.

2. 边界上给定法向速度 $\frac{\partial\varphi}{\partial n}$(这种情形下，$\varphi$ 可相差一常数).

3. 在一部分边界上给定 φ，而在另一部分边界上给定 $\frac{\partial\varphi}{\partial n}$.

在平面双连通区域中，在上述三类边界条件下，如还给定速度环量条件，速度势 φ 也可被惟一确定. 在无界区域中，为保证惟一性，除上述有界区域所要求的条件外，还应加上流量条件或给定无穷远处速度的渐近行为. 根据调和函数的性质，如果 φ 在全空间满足拉普拉斯方程，并且不等于常数，则一定存在奇点. 最简单的一种孤立奇点称为点源，相应的速度势满足方程

$$\nabla^2\varphi=\delta(p-q),$$

式中 q 是某定点，是点源所在的点，p 是区域中的动

点，$p-q$ 表示 p 点和 q 点的坐标差，$\delta(p-q)$ 是狄喇克 δ. 在平面流动中，点源速度势是

$$\varphi(p)=\frac{1}{2\pi}\ln r(p,q).$$

而在空间流动中，则为

$$\varphi(p)=-\frac{1}{4\pi}\frac{1}{r(p,q)},$$

$r(p,q)$ 是 p 点到 q 点的距离.

复位势(complex potential)　流体静力学的基本概念之一. 指不可压缩流体平面无旋运动中，由速度势和流函数组成的解析函数. 在不可压缩流体的平面无旋运动中，速度势 φ 与流函数 ψ 间满足柯西-黎曼条件：

$$\frac{\partial\varphi}{\partial x}=\frac{\partial\psi}{\partial y},\quad \frac{\partial\varphi}{\partial y}=-\frac{\partial\psi}{\partial x},$$

因此，复函数

$$F(x,y)=\varphi(x,y)+\mathrm{i}\psi(x,y)$$

是复变量 $z=x+\mathrm{i}y$ 的解析函数. 函数 F 的导数

$$\frac{\mathrm{d}F}{\mathrm{d}z}=u-\mathrm{i}v$$

称为共轭复速度，它也是解析函数. 共轭复速度沿某封闭曲线 l 的积分(沿反时针方向)为

$$\oint_l\frac{\mathrm{d}F}{\mathrm{d}z}\mathrm{d}l=\Gamma+\mathrm{i}Q,$$

Γ 是沿曲线 l 的速度环量，Q 是从曲线 l 所围区域中流出的体积流量. 当 l 所围的区域中无奇点，并且此区域为单连通时，Γ 和 Q 都等于零.

叠加原理(principle of superposition)　不可压缩流体的无旋运动中的一条原则. 这条原理是指：当在不可压缩流体的无旋运动的条件下，速度势应满足运动方程(即拉普拉斯方程)，如果 $\varphi_1,\varphi_2,\cdots,\varphi_n$ 是该方程的解，则它们的线性组合也是该方程的解.

不可压缩流体无旋运动中的许多计算方法，都是以叠加原理为基础的，根据惟一性定理，只要满足边界条件，任何一种叠加形式都可以是问题的解. 但为了满足边界条件，所能采取的叠加形式不是完全任意的. 利用势论中的格林公式，可以找到在一般情形下都可适用的几种叠加形式：如果 $\psi(p,q)$ 是位于区域边界 A 上 q 点处的点源所诱导的速度势(参见“不可压缩流体的无旋运动”)，则区域中任一点 p 上的速度势可表示成下述三种形式之一：

$$\varphi(p)=\oint_A\left(\frac{\partial\varphi}{\partial n}-\varphi\frac{\partial}{\partial n}\psi(p,q)\right)\mathrm{d}A,$$

$$\varphi(p)=\oint_A\sigma(q)\psi(p,q)\mathrm{d}A,$$

$$\varphi(p)=\int m(q)\frac{\partial}{\partial n}\psi(p,q)\mathrm{d}A,$$

$\frac{\partial}{\partial n}$ 表示沿曲面 A 的外法线方向的导数. 在第一个式子中，边界上的 φ 和 $\frac{\partial\varphi}{\partial n}$ 不是彼此独立的，具体问题中只能给定其中的一个；第二个和第三个式子中的 $\sigma(q)$ 和 $m(q)$ 是边界 A 上的分布函数，它们相当于本条目开头所说的线性组合中的系数.

上述关于速度势的三种积分表示形式，在流体力学的边界元理论中比较重要.

对应于散度场和涡量场的速度场(velocity distribution with specified rate of divergence and vorticity)　根据给定的速度场，通过微分运算确定散度场和涡量场. 许多流体力学问题常常需要根据给定的散度场和涡量场确定速度场. 这一问题相当于在某区域 τ 中求解方程组

$$\begin{cases}\nabla\cdot\boldsymbol{v}=\Theta,\\ \nabla\times\boldsymbol{v}=\boldsymbol{\Omega},\end{cases}$$

这里 Θ 和 $\boldsymbol{\Omega}$ 分别是散度场和涡量场. 上述方程组有如下形式的解：

$$\begin{aligned}\boldsymbol{v}(p)=&\frac{1}{4\pi}\oint(\boldsymbol{v}\times\boldsymbol{n})\times\underset{(q)}{\mathrm{grad}}\frac{1}{r(p,q)}\mathrm{d}A\\&-\frac{1}{4\pi}\oint_A(\boldsymbol{v}\cdot\boldsymbol{n})\underset{(q)}{\mathrm{grad}}\frac{1}{r(p,q)}\mathrm{d}A\\&+\frac{1}{4\pi}\int_\tau\Theta\underset{(q)}{\mathrm{grad}}\frac{1}{r(p,q)}\mathrm{d}\tau\\&+\frac{1}{4\pi}\int_\tau\boldsymbol{\Omega}\underset{(q)}{\mathrm{grad}}\frac{1}{r(p,q)}\mathrm{d}\tau,\end{aligned}$$

式中 $\boldsymbol{n}$ 是区域 τ 的边界 A 上的单位外法线向量，$r(p,q)$ 是边界上的 q 点到区域中 p 点的距离，$\underset{(q)}{\mathrm{grad}}$ 则是以 q 点的坐标为变量的梯度算子. 如果 $\boldsymbol{\Omega}$ 和 Θ 都恒等于零，从上式立即得到不可压缩流体无旋运动解的一种形式. 对具体问题来说，边界上的 $\boldsymbol{v}\times\boldsymbol{n}$ 和 $\boldsymbol{v}\cdot\boldsymbol{n}$ 不是彼此独立的.

毕奥-萨伐尔定理(Biot-Savart theorem)　关于无界的不可压缩流体中孤立涡线所诱导的速度场的一条定理：无界空间中如有一孤立涡管，涡管外的流动是无旋的，涡管内的涡量分布为 $\boldsymbol{\Omega}$，当流体是不可压缩时，任意点 p 处的速度(参见“对应于散度场和涡量场的速度场”)为

$$\boldsymbol{v}(p)=\frac{1}{4\pi}\int_\tau\boldsymbol{\Omega}\times\mathrm{grad}\frac{1}{r(p,q)}\mathrm{d}\tau,$$

τ 是涡管所围区域. 如果涡管截面的线尺寸远小于空间点 p 到涡管的最小距离时，可认为此涡管是一有有限强度的集中涡线，上式可写成

$$\boldsymbol{v}(p)=\frac{\Gamma}{4\pi}\int\frac{\boldsymbol{r}(p,q)}{r^3(p,q)}\times\mathrm{d}\boldsymbol{l},$$

Γ 是涡管即集中涡线的强度，$\boldsymbol{r}(p,q)$ 是涡线上的点 q 到空间点 p 的向径，

$$\boldsymbol{r}(p,q)=\boldsymbol{r}(p)-\boldsymbol{r}(q).$$

流体动力学(fluid dynamics)　流体力学的一个分支学科. 它研究作为连续介质的流体在外力作

用下的运动规律及它与其他运动形态之间的相互作用. 流体动力学的基本内容包括:流体力学基本方程组,涡量场的动力学性质,在各种为简化问题而建立的物性模型和运动学模型基础上的流动规律,流体力学方程组在一些特殊条件下的简化及有关的求解方法.

流体力学基本方程组(basic equations of fluid mechanics)　流体运动时所要遵循的基本物理定律的数学表示式. 最基本的物理定律是:

1. 质量守恒. 任一部分流体的质量在流动过程中保持不变. 反映这一规律的方程称为连续性方程. 在无质量扩散的情况下,其微分形式是

$$\frac{\partial\rho}{\partial t}+\operatorname{div}(\rho\boldsymbol{v})=0,$$

或

$$\frac{\mathrm{D}\rho}{\mathrm{D}t}+\operatorname{div}\boldsymbol{v}=0,$$

这里 ρ 是流体密度,$\boldsymbol{v}$ 是速度,$\mathrm{D}/\mathrm{D}t$ 是随体导数(参见"随体导数"). 对不可压缩流体,$\operatorname{div}\boldsymbol{v}=0$,因此有 $\mathrm{D}\rho/\mathrm{D}t=0$,即流体密度在流动过程中保持不变.

2. 动量方程. 任一部分流体的动量增加率等于包括来自周围流体对它作用在内的所有外力之和. 反映这一规律的方程称为动量方程,其微分形式是

$$\rho\frac{\mathrm{D}u_i}{\mathrm{D}t}=-\frac{\partial p}{\partial x_i}+\frac{\partial\tau_{ji}}{\partial x_j}+\rho f_i,$$

式中 x 是直角坐标,τ_{ji} 是粘性应力张量在此直角坐标系下的各分量,u_i 和 f_i 是速度分量和质量力分量,p 和 ρ 是流体的压力和密度. 式中已用了约定求和记号. 对粘性系数为常数的不可压缩牛顿流体,动量方程可写成

$$\rho\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}=-\nabla p+\mu\nabla^2\boldsymbol{v}+\rho\boldsymbol{f},$$

此方程组称为纳维-斯托克斯方程. 对理想流体,动量方程可进一步简化为

$$\rho\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}=-\nabla p+\rho\boldsymbol{f},$$

此方程组称为欧拉方程.

3. 能量方程. 流动进程中任一部分流体的总能量(内能与动能之和)的增加率,等于包括周围流体在内的外界在单位时间内对它做的功和给予的热量之和. 反映这一规律的方程称为能量方程,其微分形式是

$$\rho\frac{\mathrm{D}e}{\mathrm{D}t}=\rho q+\frac{\partial}{\partial x_j}\left(k\frac{\partial T}{\partial x_j}\right)-p\frac{\partial u_i}{\partial x_i}+\tau_{ij}\frac{\partial u_i}{\partial x_j},$$

式中 e 是单位质量流体的内能,q 为单位时间内由热源给单位质量流体的热量,k 是热传导系数,T 是温度. 能量方程中比较复杂的项是

$$\tau_{ij}\frac{\partial u_i}{\partial x_j},$$

它是单位体积中流体的内摩擦力在单位时间内所做的功,这部分功将不可逆地转化为热,通常称

$$\tau_{ij}\frac{\partial u_i}{\partial x_j}$$

为耗散函数. 对理想流体,粘性和热传导都不存在,能量方程可简化为

$$\rho\frac{\mathrm{D}e}{\mathrm{D}t}=\rho q-p\frac{\partial u_i}{\partial x_i}\text{ 或 }T\frac{\mathrm{D}S}{\mathrm{D}t}=q,$$

S 是单位质量流体的熵. 此式与可逆过程中的热力学关系式相同.

纳维-斯托克斯方程(Navier-Stokes equation)　即不可压缩粘性流体的动量方程(参见"流体力学基本方程组"). 由法国数学家、工程师纳维(Navier,C.(-L.-M.-H.))于 1821 年和英国物理学家、数学家斯托克斯(Stokes,G. G.)于 1845 年分别提出而得名.

欧拉方程(Euler equation)　即理想流体的动量方程. 由瑞士数学家欧拉(Euler,L.)于 1755 年得到而命名(参见"流体力学基本方程组").

耗散函数(dissipation function)　见"流体力学基本方程组".

涡量方程(equation of vorticity)　从流体力学基本方程组直接导出的涡量所满足的方程. 有以下几种重要情形:

1. 如果流体密度和粘性系数都是常数,质量力有势,涡量 $\boldsymbol{\Omega}$ 满足方程

$$\frac{\mathrm{D}\boldsymbol{\Omega}}{\mathrm{D}t}=(\boldsymbol{\Omega}\cdot\nabla)\boldsymbol{v}+\nu\nabla^2\boldsymbol{\Omega},$$

式中 $\boldsymbol{v}$ 是速度,ν 是运动粘性系数,它等于流体粘性系数与流体密度之比.

2. 如果流体的粘性作用可以不计,质量力有势,涡量方程有如下的形式:

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\boldsymbol{\Omega}}{\rho}\right)=\left(\frac{\boldsymbol{\Omega}}{\rho}\cdot\nabla\right)\boldsymbol{v}+\frac{1}{\rho^3}(\operatorname{grad}\rho\times\operatorname{grad}p),$$

这里 p 和 ρ 是流体的压力和密度. 可以看出,如果流动是无旋的,则密度 ρ 应可以表示成压力 p 的函数,即流体应当是正压的;反过来,如果流体非正压(斜压),流动一定是有旋的.

3. 正压流体在有势质量力作用下的平面或轴对称理想流动中,涡量方程可简化为

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\boldsymbol{\Omega}}{r^\sigma}\right)=0,$$

指数 σ 在平面流动中为 0,在轴对称流动中为 1,r 是空间点到对称轴的距离. 从上式可见,如果流体是不可压缩的,并且无质量扩散,则在定常流动的条件下,Ω/r^σ 将是流函数 ψ 的函数.

开尔文定理(Kelvin theorem)　流体力学中的一条基本定理. 它指出:在有势质量力作用下,理想

正压流体中,沿任一由流体质点所组成的封闭曲线的速度环量在流动过程中保持不变.由开尔文定理可推出:在有势质量力作用下,如果理想正压流体中某一部分流体的运动在某时刻是无旋的,则这部分流体的运动在此时刻以前或以后的任一时刻都是无旋的;反之,若某时刻这部分流体的运动有旋,则在此时刻以前或以后的任何时刻,这部分流体的运动是有旋的.从涡量方程(参见"涡量方程")也可推出由开尔文定理得出的上述推论.开尔文定理指出了无旋运动理论的实用背景.

亥姆霍兹定理(Helmholtz theorem) 关于涡管动力学性质的一条定理.此定理指出:在有势质量力作用下,理想正压流体中组成涡线、涡面和涡管的流体质点将永远组成涡线、涡面和涡管,并且涡管的强度在运动过程中保持不变.

音速(speed of sound) 流体力学中的物理量.指声音在介质中的传播速度.音速也是微小压力扰动在介质中传播的速度.音速的计算公式是

$$a=\sqrt{\left(\frac{\partial p}{\partial \rho}\right)_S},$$

式中$\left(\frac{\partial p}{\partial \rho}\right)_S$表示在等熵条件下压力随密度的变化率.介质中不同的点,可以有不同的音速.当流速大于音速时,流体中的小扰动不能向上游传播,一个指定点上的流动状态只受某特定区域中扰动的影响.在数学上表明:当流速大于音速时,相应的方程是双曲型的.

马赫数(Mach number) 表征流动的常数.是流速与当地音速之比.马赫数 $Ma=v/a$,v 为流速,a 为音速.马赫数处处大于 1 的流动称为超音速流动,处处小于 1 的流动称为亚音速流动,其余称为跨音速流动.马赫数由奥地利学者马赫(Mach,E.)首先引进而得名.

速度势方程(velocity potential equation) 速度势 φ 所满足的方程.一般情况下,速度势 φ 所满足的方程是

$$\left(1-\frac{u^2}{a^2}\right)\frac{\partial^2\varphi}{\partial x^2}+\left(1-\frac{v^2}{a^2}\right)\frac{\partial^2\varphi}{\partial y^2}+\left(1-\frac{w^2}{a^2}\right)\frac{\partial^2\varphi}{\partial z^2}$$
$$-\frac{1}{a^2}\frac{\partial^2\varphi}{\partial t^2}-2\frac{uv}{a^2}\frac{\partial^2\varphi}{\partial x\partial y}-2\frac{vw}{a^2}\frac{\partial^2\varphi}{\partial y\partial z}-2\frac{wu}{a^2}\frac{\partial^2\varphi}{\partial z\partial x}$$
$$-2\frac{u}{a^2}\frac{\partial^2\varphi}{\partial x\partial t}-2\frac{v}{a^2}\frac{\partial^2\varphi}{\partial y\partial t}-2\frac{w}{a^2}\frac{\partial^2\varphi}{\partial z\partial t}=0,$$

x,y,z 为某取定的直角坐标系三个坐标分量,u,v,w 是相应的三个速度分量,a 是音速,t 是时间.在不可压缩流体中,音速趋于无穷,速度势方程简化为拉普拉斯方程

$$\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}+\frac{\partial^2\varphi}{\partial z^2}=0.$$

伯努利方程(Bernoulli equation) 流体力学的基本方程.指在有势质量力作用下,理想正压流体运动中,以有限形式表示的能量方程.伯努利方程有如下几种重要形式:

1. 在定常流动中,沿流线时,

$$\Pi+\frac{v^2}{2}+U=c(\psi),$$

式中 v 为流速,U 为质量力的势函数,Π 是压力函数,$c(\psi)$是一个沿流线不变的常数,在不同的流线上,它可以取不同的值.

2. 在无旋运动中和在整个流场中,

$$\frac{\partial\varphi}{\partial t}+\Pi+\frac{v^2}{2}+U=0,$$

φ 为速度势.在定常流动中,应取$\frac{\partial\varphi}{\partial t}$为常数.

3. 如果绝对运动是无旋的,在运动坐标系中,伯努利方程可写成

$$\frac{\partial'\varphi}{\partial t}+\Pi+\frac{v_r^2}{2}-\frac{v_e^2}{2}+U=0,$$

$\frac{\partial'\varphi}{\partial t}$是运动坐标系中的速度势相对导数,$v_r$ 是相对速度,v_e 是牵连速度.

4. 在不可压缩均质流体的平面或轴对称定常流动中,

$$\frac{p}{\rho}+\frac{v^2}{2}+U+F(\psi)=c,$$

式中 c 是常数,$F(\psi)$是流函数的函数,它与涡量Ω间的关系是

$$F(\psi)=\frac{\Omega}{r^{\sigma}},$$

指数 σ 在平面流动中取零,在轴对称流动中取 1,r 是空间点到对称轴的距离.

托里切利公式(Torricelli formula) 流体力学基本公式.是关于大容器中的水在重力作用下从小孔流出的水速公式.托里切利公式的一般形式是

$$v=\sqrt{2gh},$$

h 为小孔到水平面的垂直距离,g 是重力加速度.托里切利公式指出,小孔处的水速与质点从水面自由降落到小孔时的速度相同.托里切利公式可从伯努利方程推出.在实际流体中由于存在粘性阻力,射流速度要小一些,对圆孔出流,射流速度约为上述公式计算值的 0.98 倍.

小扰动理论(small-perturbation theory) 分析和研究某些流体力学问题的一种近似理论.小扰动的含义是流场中置入一物体,或其他原因使原有速度场有所改变,但改变值与未改变时的值相比是很小的.如果未扰流场是均匀的,流速方向平行于 x 轴,其大小为 u_∞,受到扰动以后,x,y,z 方向的速度分量将改变为 $u_\infty+u$,v 和 w,在小扰动情形下有

$$\left|\frac{u}{u_\infty}\right| \ll 1, \left|\frac{v}{u_\infty}\right| \ll 1, \left|\frac{w}{u_\infty}\right| \ll 1.$$

在无旋运动中，如果流动是定常的，u_∞不太接近于未扰流场中的音速 a_∞，并且马赫数 $Ma_\infty = u_\infty/a_\infty$并不很大，速度势方程可简化为

$$(1-Ma_\infty^2)\frac{\partial^2 \varphi}{\partial x^2}+\frac{\partial^2 \varphi}{\partial y^2}+\frac{\partial^2 \varphi}{\partial z^2}=0.$$

可见，在亚音速流动中，速度势方程是椭圆型的，并且只需作简单的变换即可将此方程变为拉普拉斯方程，即亚音速流动与不可压缩流动之间一定存在某种相似关系. 在超音速流动中，速度势方程是双曲型的，并且是一个简单的波动方程. 在跨音速流动中，速度势方程不能线性化，但可以简化为

$$(1-Ma_\infty^2)\frac{\partial^2 \varphi}{\partial x^2}+\frac{\partial^2 \varphi}{\partial y^2}+\frac{\partial^2 \varphi}{\partial z^2} = (\gamma+1)\frac{Ma_\infty}{a_\infty}\frac{\partial \varphi}{\partial x}\frac{\partial^2 \varphi}{\partial x^2},$$

这里已认为流体是比热为常数的完全气体，γ 是定压比热与定容比热之比. 跨音速情形下的速度势方程是混合型方程，在一些区域中，它是椭圆型的，在另一些区域中，它是双曲型的.

升力(lift)　一种形式的力. 指物体在静止流体中运动时受到的与运动方向垂直的分力. 在航空问题中，物体运动往往是沿水平方向的，因此如果取一坐标系 $Oxyz$，让 z 轴沿铅直向上的方向，x 轴和 y 轴在水平平面内，并且让 x 轴指向物体运动的反方向，则也可把流体作用在物体上的合力在 z 轴上的投影定义为升力，而把 y 方向上的分力称为侧向力. 如果物体是作等速直线运动的，并且不考虑流体的可压缩性和粘性，同时忽略质量力的作用，则升力可用下式计算：

$$L = \rho u_\infty \int_{y_1}^{y_2} \Gamma \mathrm{d}y,$$

式中 ρ 是流体密度，u_∞是物体运动的速度，Γ 是速度环量

$$\Gamma = \oint_l u\mathrm{d}x + w\mathrm{d}z,$$

u 和 w 是流体在 x 方向和 y 方向的速度分量，l 是与 y 轴垂直的平面内的且围绕着物体的封闭曲线，线积分的走向是：当逆着 y 轴的方向去看时是反时针方向的. 升力公式中的积分上下限 y_1 和 y_2 是物体上两个端点的 y 坐标. 在空间流动中，根据开尔文定理，此问题中沿任一封闭曲线的速度环量等于零，因而物体的升力也等于零. 但是如果考虑物体下游有尾迹存在，并且把它简单地考虑为一个压力连续的切向间断面，则穿过尾迹的封闭曲线上的速度环量将不等于零，而等于尾迹上速度势的间断值，物体相应地将有升力. 在平面流动中，由于包围物体的封闭曲线 l 所围的区域是双连通区域，环量可以不等于零，升力公式将简单地变为 $L=\rho u_\infty \Gamma$，此关系式称为茹柯夫斯基公式.

在流体力学中，确定环量 Γ 是一件困难的事. 对尖后缘物体，可假定尖后缘处速度有限，由此去确定速度环量，这一假定称为库塔-茹柯夫斯基假定. 对非尖后缘物体，在理想流动的条件下尚无有效办法确定速度环量.

茹柯夫斯基公式(Joukowski formula)　见“升力”.

库塔-茹柯夫斯基假定(Kutta-Joukowski hypothesis)　见“升力”.

奥斯瓦梯契公式(Oswatitsch formula)　定常绕流中关于物体阻力的公式. 在均匀静止流体中，物体以速度 v 作等速直线运动时，流体作用在物体上的合力在物体运动方向上的投影称为物体的阻力. 如果把参考系取在物体上，物体在原静止的流体中作等速直线运动的问题，将替换为定常绕流问题. 假定能找到一个足够大的封闭曲面 A，它包围物体，并把问题中所有的能量耗散过程都包在里面，即在此曲面附近，流体的速度和温度等物理量的梯度，小到使曲面上的粘性力和热通量都可以完全忽略不计，则有

$$vD - H = T_\infty \oint_A \rho S v_n \mathrm{d}A,$$

式中 D 是物体阻力，H 是单位时间内流体传给物体的热量，T_∞是来流温度，S 是单位质量流体的熵，ρ 是流体密度，v_n 是曲面 A 上的流体在外法线方向上的速度分量. 上式还可以写成

$$vD - H = T_\infty \oint_A \rho \frac{\mathrm{D}S}{\mathrm{D}t}\mathrm{d}A.$$

在等温流动中，如果不存在热源，奥斯瓦梯契公式还可写成

$$vD = \oint_\tau \varphi \mathrm{d}\tau,$$

式中 φ 是耗散函数，τ 是曲面 A 和物面间的流动区域. 在理想流动中，耗散函数 φ 等于零，因此从上式可得出：物体在静止的理想流体中作等速直线运动时，阻力将等于零. 这一结论称为达朗贝尔佯谬.

达朗贝尔佯谬(d'Alembert paradox)　见“奥斯瓦梯契公式”.

布拉休斯定理(Blasius theorem)　流体动力学的一条定理. 是不可压缩流体绕物体的平面定常无旋运动中关于流体作用在物体上的合力和合力矩的定理. 布拉休斯定理指出：如果 $F(z)$是复位势，X 和 Y 是合力在 x 轴和 y 轴上的投影，M 和 N 是对原点的合力矩在 x 轴和 y 轴上的投影，ρ 为密度，则有

$$X - \mathrm{i}Y = \frac{\mathrm{i}}{2}\rho \oint \left(\frac{\mathrm{d}F}{\mathrm{d}z}\right)^2 \mathrm{d}z,$$

$$M + \mathrm{i}N = -\frac{\rho}{2}\oint z\left(\frac{\mathrm{d}F}{\mathrm{d}z}\right)^2 \mathrm{d}z,$$

积分沿反时针方向包围物体的封闭曲线进行.

卡门涡街(Kármán vortex street) 流体动力学概念. 指在一定条件下,绕流问题中从物体两侧周期性地脱落的旋转方向相反的涡旋组成的两列有规则的涡列. 卡门涡街的性态与雷诺数有关. 当雷诺数为 50～300 时,涡旋脱落有周期性规律,并且在某种特殊条件下涡街是稳定的. 当雷诺数大于 300 时,涡旋脱落开始表现出随机性.

液体自由表面波(free surface wave in liquid) 液体自由面上的波浪运动. 如果不考虑表面张力的作用,并且认为表面波是因重力作用而引起的,则这类问题的数学提法是: $\nabla^2\varphi = 0$. 在自由面 $z = \zeta(x, y, t)$ 上,

$$\frac{\partial\zeta}{\partial t} + \frac{\partial\varphi}{\partial x}\frac{\partial\zeta}{\partial x} + \frac{\partial\varphi}{\partial y}\frac{\partial\zeta}{\partial y} = \frac{\partial\varphi}{\partial z},$$

$$\frac{\partial\varphi}{\partial t} + \frac{1}{2}\left[\left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2 + \left(\frac{\partial\varphi}{\partial z}\right)^2\right] + g\zeta = 0;$$

在槽底(河床) $z = -h_0(x, y)$ 上,

$$\frac{\partial\varphi}{\partial z} + \frac{\partial\varphi}{\partial x}\frac{\partial h_0}{\partial x} + \frac{\partial\varphi}{\partial y}\frac{\partial h_0}{\partial y} = 0.$$

这里坐标原点取在无波浪时的自由面上, φ 是速度势, ζ 是自由面的位移, g 是重力加速度,并且假定自下而上大气压均匀分布. 这是一个非线性问题. 在波幅很小的情况下,此问题可线性化. 这时, $z = \zeta(x, y, t)$ 上的条件可简化为

$$\frac{\partial\zeta}{\partial t} = \frac{\partial\varphi}{\partial z}, \quad \frac{\partial\varphi}{\partial t} + g\zeta = 0.$$

在这种情形下,如果水深 h_0 为常数,可得到通解

$$\varphi = -g\mathrm{i}\iint_{-\infty}^{+\infty} \frac{\cos h\boldsymbol{k}(z + h_0)}{\omega \cos h(\boldsymbol{k}h_0)} \cdot F_1(k_1, k_2)\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{R} - \omega t)}\mathrm{d}k_1\mathrm{d}k_2 + g\mathrm{i}\iint_{-\infty}^{+\infty} \frac{\cos h\boldsymbol{k}(z + h_0)}{\omega \cos h(\boldsymbol{k}h_0)} \cdot F_2(k_1, k_2)\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{R} + \omega t)}\mathrm{d}k_1\mathrm{d}k_2,$$

式中 $\mathrm{i} = \sqrt{-1}$,

$$\boldsymbol{k} = (k_1, k_2), \ \boldsymbol{R} = (x, y), \ k = \sqrt{k_1^2 + k_2^2},$$

k 和 ω 分别称为波数和角频率, k 与 ω 之间有

$$\omega = \sqrt{gk \tan h(kh_0)},$$

此关系式称为色散关系. $F_1(k_1, k_2)$ 和 $F_2(k_1, k_2)$ 由初始条件决定. 对水深 h_0 为常数的情形,线性表面波问题可归结为如下的变分问题

$$\delta\iiint_R L\mathrm{d}x\mathrm{d}y\mathrm{d}t = 0,$$

$$L = -\rho\int_{h_0}^{\zeta}\left[\frac{\partial\varphi}{\partial t} + \frac{1}{2}(\nabla\varphi)^2 + gz\right]\mathrm{d}z,$$

R 为 (x, y, t) 空间中的区域.

浅水理论(shallow water theory) 水力系的一个分支. 波长远大于水深的流体动力学理论. 如果静止时的水深 h_0 为常数,不考虑与地转有关的科氏力的作用,则浅水问题中的控制方程是

$$\frac{\partial h}{\partial t} + \frac{\partial(hx)}{\partial t} + \frac{\partial(hy)}{\partial t} = 0,$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -g\frac{\partial h}{\partial x},$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -g\frac{\partial h}{\partial y},$$

这里 h 是从自由面量起的水深, $h = h_0 + \zeta$, ζ 是自由面位移, g 是重力加速度. 平面流动的情形将变得比较简单,而且所有物理量对 y 的导数及速度分量 v 都等于零. 如果波是沿着 x 轴方向向右传到水深为 h_0 的未扰水域的,自由面位移将满足

$$\frac{\partial\zeta}{\partial t} + \left[3\sqrt{g(h_0 + \zeta)} - 2\sqrt{gh_0}\right]\frac{\partial\zeta}{\partial x} + \frac{1}{6}\sqrt{gh_0}h_0^2\frac{\partial^3\zeta}{\partial x^3} = 0.$$

如果令

$$c_0 = \sqrt{gh_0}, \quad \alpha = \frac{c_0 h_0^2}{6},$$

上述自由面方程可写成

$$\frac{\partial\zeta}{\partial t} + c_0\left(1 + \frac{3}{2}\frac{\zeta}{h_0}\right)\frac{\partial\zeta}{\partial x} + \alpha\frac{\partial^3\zeta}{\partial x^3} = 0.$$

此方程称为 KdV 方程,在孤立子理论中这是一个基本方程.

KdV 方程(Korteweg-de Vries equation) 见"浅水理论".

层流(laminar flow) 亦称片流. 流体的一种运动形式. 指粘性流体的层状运动. 在层流中,任意两个流体质点,只要此二质点足够接近,在流动过程中,这两点之间的偏离可以足够小. 在层流中,流体质点的运动是有序的,相邻两部分流体间的动量交换、能量交换和质量交换是由分子的相互作用引起的. 层流只出现在雷诺数较小的情形中,当雷诺数很大,大到超过某一临界值时,层流将失去稳定而逐渐转变为湍流.

流体运动稳定性(stability of fluid motion) 流体力学的基本概念之一. 指流动受到某初始扰动后恢复到原来状态的能力. 如果流动能恢复到原来状态,则称流动是稳定的,否则称流动是不稳定的. 流动稳定性理论研究流动稳定的条件和失稳后流动的发展变化,包括从层流转捩为湍流的过程. 在理论研究中,主要考察扰动量的发展. 如果假定扰动量很小,以致这些量所遵守的方程可以线性化,在此基础上建立的稳定性理论称为线性稳定性理论;如果扰动量不是很小,相应的方程不能线性化,这种稳定性理论称为非线性稳定性理论. 平行流的线性稳定性

是一个典型问题. 例如在两平行平板间的平面流动中,若把扰动流函数记为 $\varphi(y)e^{i\alpha(x-ct)}$,则问题将归结为研究方程

$$\left(\frac{d^2}{dy^2}-\alpha^2\right)^2\varphi = i\alpha Re\left[(u-c)\left(\frac{d^2}{dy^2}-\alpha^2\right)\varphi-\left(\frac{d^2u}{dy^2}\right)\varphi\right]$$

在条件 $y=\pm1$ 时,

$$\varphi=\frac{d\varphi}{dy}=0$$

的特征值 c 的分布. 式中 u 是速度,Re 是雷诺数,α 为扰动波数. 对给定的 Re,如果对一切 α,c 的虚部都小于零,则扰动会衰减,流动是稳定的;如果某些 c 的虚部大于零,流动是不稳定的. 上述方程称为奥尔-索末菲方程. 平面层流的边界层流动中的稳定性也可用奥尔-索末菲方程进行分析.

奥尔-索末菲方程(Orr-Sommerfeld equation) 见"流体运动稳定性".

湍流(turbulent flow) 流体的一种运动形式. 指速度、压力等物理量在空间和时间中都发生不规则脉动的流体运动. 湍流是在雷诺数比较大的情形下发生的. 在湍流中,由于流体微团的随机脉动,流体内部的动量、能量和质量输运能力比层流的情形要高出若干个量级,因此湍流运动中表现的粘性系数、热传导系数和质量扩散系数比单纯由分子作用产生的相应系数要大得多. 在湍流研究中,一般地,纳维-斯托克斯方程仍有效,核心问题是求该方程的统计解. 因为方程的非线性性质和湍流解的不规则性,湍流是流体力学,乃至整个物理学领域中最为困难的问题,目前还没有成熟的精确理论. 目前,湍流理论中有半经验理论和统计理论,只有在相当特殊的条件下,这些理论才能得到比较符合实际的结果.

雷诺方程(Reynold equation) 湍流理论中的基本方程. 雷诺方程是从流体力学基本方程组取统计平均后得到的. 在粘性系数 μ 和流体密度 ρ 都为常数的情况下,对纳维-斯托克斯方程作统计平均,可得到

$$\frac{\partial\overline{u}_i}{\partial t}+\overline{u}_\alpha\frac{\partial\overline{u}_i}{\partial x_\alpha}=-\frac{1}{\rho}\frac{\partial\overline{p}}{\partial x_i}+\nu\frac{\partial^2\overline{u}_i}{\partial x_\alpha\partial x_\alpha}-\frac{\partial}{\partial x_\alpha}\overline{u_i'u_\alpha'},$$

这里 $\overline{u}_i$,$\overline{p}$ 是统计平均速度和统计平均压力,u_i' 是脉动速度,ν 是流体的运动粘性系数,它等于流体的粘性系数与流体密度之比. 由于湍流脉动而产生的附加应力张量 $-\rho\overline{u_i'u_\alpha'}$ 称为雷诺应力张量或湍流应力张量. 雷诺方程和连续性方程合在一起,并不能构成封闭方程组. 在湍流理论中,一个根本的也是最困难的工作,是建立控制湍流运动的封闭方程组.

湍流半经验理论(semi-empirical theorem in turbulent flow) 描述湍流的一种理论. 是根据实验结果和物理方面的某些考虑,使湍流平均运动方程封闭并做出有关分析和计算的理论. 在固壁附近,在平面射流问题中,在直管中,湍流运动中雷诺应力张量的主要项是 $-\rho\overline{u'v'}$,ρ 是流体密度,u' 是流动方向上的脉动速度,v' 是与流动方向垂直的脉动速度. 最早的湍流半经验理论中把此雷诺应力表示成

$$-\rho\overline{u'v'}=\rho l^2\left|\frac{d\overline{u}}{dy}\right|\frac{d\overline{u}}{dy},$$

$\overline{u}$ 是平均速度,l 是一个具有长度量纲的量,称为混合长. 在固壁附近,可认为 l 是与到固壁的距离成正比的;在自由射流问题中,l 是与射流宽度成正比的. 有关的比例系数是通过实验确定的. 这一半经验理论称为混合长理论. 作为混合长理论的改进和发展,在湍流半经验理论中有一方程模式、两方程模式和剪应力模式. 这些模式理论中比较受到重视并已得到较多应用的是属于两方程模式的 K-ε 模式,这一模式中,将雷诺应力表示成

$$-\rho\overline{u_i'u_j'}=\mu_t\left(\frac{\partial\overline{u}_i}{\partial x_j}+\frac{\partial\overline{u}_j}{\partial x_i}\right)-\frac{2}{3}\rho K\delta_{ij},$$

式中 K 是脉动能,$K=\overline{u_i'u_j'}/2$,μ_t 是表观湍流粘性系数,μ_t 可表示成 $\mu_t=c_1\rho K^2/\varepsilon$,$\varepsilon$ 为湍流脉动耗散函数,

$$\varepsilon=\nu\overline{\frac{\partial u_i'}{\partial x_\alpha}\frac{\partial u_i'}{\partial x_\alpha}},$$

K 和 ε 满足方程组

$$\frac{DK}{Dt}=\nabla\cdot\left(c_2\frac{\mu_t}{\rho}\nabla K\right)-G-\varepsilon,$$

$$\frac{D\varepsilon}{Dt}=\nabla\cdot\left(c_3\frac{\mu_t}{\rho}\nabla\varepsilon\right)+c_4\frac{\varepsilon}{K}G-c_5\frac{\varepsilon^2}{K}.$$

式中

$$G=\frac{\mu_t}{\rho}\left[\frac{\partial\overline{u}_i}{\partial x_j}+\frac{\partial\overline{u}_j}{\partial x_i}\right]\frac{\partial\overline{u}_i}{\partial x_j},$$

c_1,c_2,c_3,c_4,c_5 为模式常数.

混合长(mixing length) 见"湍流半经验理论".

湍流模式理论(model theory in turbulent flow) 见"湍流半经验理论".

管道流动(pipe flow) 流体的一种运动形式. 指流体在管中的流动. 流体在直管中流动时,以平均流速 v、管道直径 D 和流体运动粘性系数组成的雷诺数小于 2000 左右时,管道中的流动是层流,当雷诺数大于 4000 时,一般工业管道中的流动为湍流. 不可压缩流体在长直管道中作定常层流运动时,确定管道截面上的速度分布问题归结为求解泊松方程

$$\frac{\partial^2u}{\partial y^2}+\frac{\partial^2u}{\partial z^2}=-G,$$

式中的 y 和 z 是管道截面上的两个直角坐标,G 是常数. 如果让 x 轴指向流动方向,G 和压力 p 之间

有

$$G=-\frac{1}{\mu}\frac{\partial p}{\partial x},$$

μ 是粘性系数. 在考虑重力作用时,应将式中的 p 理解为真实压力与静水压力之差. 对圆管中的层流运动,体积流量 Q 与管道直径 D 之间有关系

$$Q=\frac{\pi}{128}GD^4,$$

此式称为哈根-泊肃叶公式. 对截面不是圆的直管,体积流量与截面线尺寸的四次方成正比的规律成立. 对湍流情形,目前还没有严格的理论分析方法,工程上常采用半经验理论去作有关计算.

斯托克斯流动(Stokes flow)　一种流动形式. 是可以把惯性力完全忽略的粘性流动. 在斯托克斯流动中,以特征速度、特征长度和流体粘性系数组成的雷诺数必须很小. 斯托克斯流动的基本方程组是

$$\begin{cases}\nabla\cdot\boldsymbol{v}=0,\\ \nabla p=\mu\nabla^2\boldsymbol{v},\end{cases}$$

这里没有考虑质量力的作用. 式中的 $\boldsymbol{v}$,p 和 μ 分别是流体速度、压力和粘性系数. 对斯托克斯流动,数学上可以作与不可压缩流体无旋运动相仿的讨论. 例如,类似于不可压缩流体无旋运动中的点源,满足方程组

$$\begin{cases}-\nabla p+\mu\nabla^2\boldsymbol{v}=\delta(x-x',y-y',z-z')\boldsymbol{e}_k,\\ \nabla\cdot\boldsymbol{v}=0\end{cases}$$

的速度场是一种基本流动. 上式中的 $\delta(x-x',y-y',z-z')$ 是狄喇克 δ,$\boldsymbol{e}_k$ 是沿第 k 个坐标轴的单位向量. 如把这一速度场记为 $\boldsymbol{w}^{(k)}$,则它的第 j 个分量是

$$w_j^{(k)}=\frac{1}{8\pi\mu}\left[\frac{1}{r}\delta_{jk}+\frac{(x_j-x_j')(x_k-x_k')}{r^3}\right],$$

r 是点 (x,y,z) 到点 (x',y',z') 的距离,δ_{jk} 是克罗内克 δ. 利用基本流动速度场 $\boldsymbol{w}^{(k)}$,区域 τ 中任意点 (x,y,z) 上的流速可写成如下的形式:

$$v_k(x_1,x_2,x_3)=\oint_A\boldsymbol{\sigma}(x_1',x_2',x_3')\cdot\boldsymbol{w}^{(k)}\mathrm{d}x_1'\mathrm{d}x_2'\mathrm{d}x_3',$$

A 是流场的边界,v_k 是沿第 k 个坐标轴的速度分量,$\boldsymbol{\sigma}$ 是一个表征强度的矢量. 上式是斯托克斯流动中边界元方法的一个基本关系式. 在更精确的小雷诺数流动分析中,斯托克斯流动可作为渐近匹配方法内部解的零级近似. 在平面绕流问题中,斯托克斯流动不存在.

斯托克斯公式(Stokes formula)　计算粘性流体对阻力的公式. 是孤立圆球在无界的粘性流体中作极缓慢移动时的阻力公式. 对半径为 a 的固体圆球,阻力 D 为

$$D=6\pi\mu av,$$

式中 μ 为粘性系数,v 为小球相对于流体的速度. 对半径为 a 的液滴,阻力为

$$D=2\pi\mu av\frac{2\mu+3\mu'}{\mu+\mu'},$$

式中 μ' 为液滴的粘性系数. 对半径为 a 的气泡,阻力为

$$D=4\pi\mu av.$$

边界层(boundary layer)　流体力学的基本概念之一. 指在大雷诺数情形下,与理想流体相邻的、其中粘性作用不能忽略的薄层. 有两类重要的边界层. 一类是附于物体表面的薄层,也称为附面层. 另一类是两股不同流束之间的过渡层. 边界层可分为层流边界层和湍流边界层. 在实际问题中,常常是一部分为层流边界层,一部分为湍流边界层,两部分之间有过渡区. 对于附于物体表面的边界层,在流体为不可压缩均质、流动为平面定常的条件下,控制方程组是

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial x}=U_eU_e'+\frac{\partial}{\partial y}\left[(\nu+\nu_t)\frac{\partial u}{\partial y}\right],$$

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0,$$

式中 y 是流场中的点到物面的距离,x 是相应垂足沿物面到物体前缘的弧长,U_e 是边界层外缘的速度,

$$U_e'=\frac{\mathrm{d}U_e}{\mathrm{d}x},$$

u 和 v 是 x 方向和 y 方向的速度分量,ν 和 ν_t 则分别是物体的和湍流表观的运动粘性系数.

对可压缩流体的情形,需要考虑粘性系数随温度的变化,同时还需要考虑能量方程. 可压缩流体的附于物体表面的定常层流边界控制方程组是

$$\begin{cases}\rho\left(u\dfrac{\partial u}{\partial x}+v\dfrac{\partial u}{\partial y}\right)=\rho_eU_eU_e'+\dfrac{\partial}{\partial y}\left(\mu\dfrac{\partial u}{\partial y}\right),\\ u\dfrac{\partial h}{\partial x}+v\dfrac{\partial h}{\partial y}=\dfrac{\partial}{\partial y}\left(k\dfrac{\partial T}{\partial y}\right)+\mu\left(\dfrac{\partial u}{\partial y}\right)^2-\rho_eU_eU_e'u,\\ \dfrac{\partial u}{\partial x}+\dfrac{\partial v}{\partial x}=0,\end{cases}$$

这里 ρ_e 是边界层外缘流体的密度,k 是热传导系数,h 是单位质量流体的焓. 在具体研究上述方程组时,还需要给出流体的状态方程、粘性系数和热传导系数随温度的变化规律.

边界层方程组的相似性解(similar solution of boundary layer equations)　边界层方程的一类解. 如果边界层中不同截面上的速度剖面 $u(x,y)$ 的差别只表现在这两截面上 x 和 y 有不同的尺度,则称这种边界层有相似性解. 在不可压缩均质流体的平面定常层流边界层流动中,当边界层外缘速度分布有幂次规律 $U_e=cx^m$ 时,相应的边界层方程组有相似性解,上式中的 c 和 m 为常数. 在这种情形下,如

令

$$\xi = x,\ \eta = \sqrt{\frac{m+1}{2}\frac{U_e x}{\nu}},\ \frac{u}{U_e} = \varphi'(\eta),$$

从边界层方程组可得到

$$\varphi''' + \varphi\varphi'' + \beta(1-\varphi'^2) = 0,\ \beta = \frac{2m}{m+1}.$$

求解上述方程的边界条件是：当 $\eta=0$ 时，$\varphi=0$，$\varphi'=0$；当 $\eta\to\infty$ 时，$\varphi'=1$.

对平板边界层，边界层外缘的速度 U_e 为常数，因此 $m=0$，$\beta=0$，相应的方程是

$$\varphi''' + \varphi\varphi'' = 0,$$

此方程称为布拉休斯方程，可用渐近方法和数值计算方法求解.

布拉休斯方程(Blasius equation)　见"边界层方层组的相似性解".

边界层方程组的变换(transformation of boundary layer equations)　边界层方程中的变量替换. 是为便于分析或揭示边界层流动的某些数学性质而进行的变量替换. 在边界层理论中，有很多变换可用来帮助讨论边界层流动. 有以下著名的三个变换：

1. 冯·米西斯变换. 取 x 和流函数 ψ 为自变量，取 $\chi=U_e^2-u^2$ 为因变量，U_e 和 u 分别为边界层外缘和边界层中的纵向速度分量，这时，不可压缩流体的平面定常层流边界层方程组可变为下述的具有标准形式的抛物型方程

$$\frac{\partial \chi}{\partial x} = \nu(U_e^2-\chi)^{1/2}\frac{\partial^2\chi}{\partial\psi^2}.$$

2. 克洛柯变换. 取 x 和纵向速度分量 u 为自变量，

$$\tau = \mu\frac{\partial u}{\partial y}$$

为因变量，则从平面定常可压缩层流边界方程可得到

$$\mu\rho u\frac{\partial}{\partial x}\left(\frac{1}{\tau}\right) + \frac{\partial^2\tau}{\partial u^2} - \mu\frac{\mathrm{d}p}{\mathrm{d}x}\frac{\partial}{\partial u}\left(\frac{1}{\tau}\right) = 0,$$

式中 $\mathrm{d}p/\mathrm{d}x$ 可用 $-\rho_e U_e U_e'$ 代替，对平板边界层有

$$\frac{\rho\mu}{\tau^2}\frac{\partial\tau}{\partial x} = \frac{1}{u}\frac{\partial^2\tau}{\partial u^2}.$$

3. 斯特瓦逊变换. 对可压缩流体的平面定常层流边界层流动，如果流体的粘性与温度有正比关系：

$$\frac{\mu}{\mu_0} = c\frac{T}{T_0},$$

μ_0 是在某参考温度 T_0 时的流体粘性系数，c 是某常数，则作变换

$$\xi = \int_0^x c\frac{a_e}{a_0}\frac{p_e}{p_0}\mathrm{d}x,\quad \eta = \frac{a_e}{a_0}\int_0^y\frac{\rho}{\rho_0}\mathrm{d}y$$

后，可把可压缩情形下的边界层方程组变到非常接近不可压缩情形下的边界层方程组的形式. 这里的 a_e 和 a_0 分别是边界层外缘的音速和绝热滞止状态下的音速，p_e 和 p_0 则是相应条件下的压力，ρ 和 ρ_0 是相应条件下的密度.

流体力学中的主要相似判据(principal criteria of similitude in fluid mechanics)　一些重要的相似判据. 主要有：

1. 雷诺数. $Re=\rho UL/\mu$，U 和 L 是给定问题中的特征速度和特征长度. ρ 和 μ 是流体的密度和粘性系数. 雷诺数表征惯性力与粘性力之比.

2. 弗劳德数. $Fr=U/\sqrt{gL}$，g 是重力加速度. 弗劳德数表征惯性力与重力之比.

3. 马赫数. $Ma=U/a$，a 是流体中的当地音速. 马赫数可用来度量流体的可压缩性对流动的影响. 当 $Ma\ll 1$ 时，比如 $Ma<0.3$ 时，流体可近似地认为是不可压缩的.

4. 普朗托数. $Pr=\mu c_p/k$，c_p 和 k 分别是流体的定压比热和热传导函数. 普朗托数可用来度量流体中动量交换和热交换的相对重要性.

5. 斯特劳哈尔数. $Sr=fL/U$，f 为流动的特征频率. 斯特劳哈尔数表征对流时间尺度与特征周期之比，当此量纲一的数很小时，可认为流动是定常的.

6. 努塞尔数. $Nu=\alpha L/k$，α 为散热系数，$\alpha=q/\Delta T$，q 为单位时间内从单位面积上传递的热量，ΔT 是温差.

7. 格拉晓夫数. $Gr=\beta_t gL^3\Delta T/\nu^2$，$\beta_t$，$\nu$ 分别是体胀系数和运动粘性系数. 在自由对流问题中，格拉晓夫数是一个重要的量纲一的量.

8. 韦伯数. $We=\rho U^2L/\alpha$，α 是表面张力系数. 韦伯数表征惯性力与表面张力之比.

激波(shock wave)　亦称冲击波. 流体力学的基本概念之一. 指介质中的压力间断面. 在超音速流动、爆炸和冲击等许多情况下都可以发生激波. 激波可认为是非线性压缩波系中后续波赶上前行波，因而在第一个波阵面附近叠加而形成的间断面. 激波前的介质相对于激波的流速方向与激波间的夹角，称为激波角. 如果激波角为直角，称此激波为正激波，否则称为斜激波. 介质通过激波时遵守质量守恒、动量守恒和能量守恒定律：

$$\begin{cases}\rho_1 v_{1n} = \rho_2 v_{2n},\\ p_1 + \rho_1 v_{1n}^2 = p_2 + \rho_2 v_{2n}^2,\\ v_{1t} = v_{2t},\\ h_1 + \dfrac{1}{2}v_{1n}^2 = h_2 + \dfrac{1}{2}v_{2n}^2,\end{cases}$$

式中 p,ρ,v,h 分别是介质的压力、密度、速度和单位质量介质的焓，下标 1 和 2 分别表示激波前、后的参量，下标 n 和 t 分别表示沿激波法向和切向的分

量. 上述方程组称为激波相容条件. 从激波相容条件可得出

$$h_2 - h_1 = \frac{1}{2}\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right)(p_2 - p_1),$$

此式称为兰金-雨果尼欧关系. 当激波强度, 即 $(p_2 - p_1)/p_1$ 很小时, 从上式可得到

$$S_2 - S_1 \approx \frac{1}{12}\frac{1}{T_1}\frac{\partial^2}{\partial p^2}\left(\frac{1}{\rho}\right)_S (p_2 - p_1)^3,$$

S 是单位质量介质的熵, T_1 是激波前的介质温度. 式中右边的下标 S 表示求导时熵应保持不变. 对比热为常数的完全气体, 利用状态方程、激波相容条件和热力学关系式可得到

$$\frac{p_2}{p_1} = 1 + \frac{2\gamma}{\gamma+1}(Ma_{1n}^2 - 1),$$

$$\frac{\rho_2}{\rho_1} = 1 + \frac{Ma_{1n}^2 - 1}{1 + \frac{\gamma-1}{2}Ma_{1n}^2},$$

$$\frac{T_2}{T_1} = 1 + \frac{2(\gamma-1)(Ma_{1n}^2 - 1)}{(\gamma+1)^2 Ma_{1n}^2}(\gamma Ma_{1n}^2 + 1),$$

$$Ma_{2n}^2 = \frac{2 + (\gamma-1)Ma_{1n}^2}{2\gamma Ma_{1n}^2 - (\gamma-1)},$$

$$Ma_2^2 = \frac{2 + (\gamma-1)Ma_{1n}^2}{2\gamma Ma_{1n}^2 - (\gamma-1)} + \frac{2Ma_{1t}^2}{(\gamma-1)Ma_{1t}^2 + 2},$$

$$\cot\theta = \tan\beta\left[\frac{(\gamma+1)Ma_{1n}^2}{2(Ma_{1n}^2 - 1)} - 1\right],$$

β 是激波角, θ 为通过激波时气流的偏转角, γ 为定压比热与定容比热之比, Ma 为马赫数, 它等于流速与当地音速之比.

冲击波(shock wave)　即"激波".

膨胀波(expansion wave)　流体力学的基本概念之一. 指流体中扰动区与未扰动区的分界面. 流体通过此界面会使压力降低. 膨胀波问题中一个典型例子是超音速气流绕凸角的定常流动, 这种流动称为普朗托-迈耶尔流动. 在这种流动中, 气流逐渐从一个方向转到另一个方向, 同时不断加速. 速度 v 的增量 $\mathrm{d}v$ 与气流转角 $\mathrm{d}\theta$ 之间有如下关系:

$$\frac{\mathrm{d}v}{v} = \frac{\mathrm{d}\theta}{\sqrt{Ma^2 - 1}},$$

或

$$\theta = \int \frac{\sqrt{Ma^2 - 1}}{1 + \frac{\gamma-1}{2}Ma^2}\frac{\mathrm{d}Ma}{Ma} + \text{常数},$$

式中 Ma 为马赫数. 在具体分析时可引进函数

$$\nu(Ma) = \int_1^{Ma} \frac{\sqrt{Ma^2 - 1}}{1 + \frac{\gamma-1}{2}Ma^2}\frac{\mathrm{d}Ma}{Ma} = \sqrt{\frac{\gamma+1}{\gamma-1}}\arctan\sqrt{\frac{\gamma-1}{\gamma+1}(Ma^2-1)} - \arctan\sqrt{Ma^2-1},$$

此函数称为普朗托-迈耶尔函数.

拉瓦尔管(Laval nozzle)　一种特别形状的管道. 是一种先收缩后扩张的管道. 通过拉瓦尔管, 气流可以从亚音速过渡到超音速. 拉瓦尔管在超音速风洞、喷气发动机、火箭推进技术、蒸汽透平等许多部门中都有广泛应用. 当拉瓦尔管进出口的压力比足够大时, 气流先在收缩段加速, 在管道最小截面处达到音速, 然后在扩张段继续加速, 其过程可用一维等熵流动的模型去估算, 对完全气体, 有

$$\frac{p}{p^*} = \left[\frac{\gamma+1}{2 + (\gamma-1)Ma^2}\right]^{\frac{\gamma}{\gamma-1}},$$

$$\frac{\rho}{\rho^*} = \left[\frac{\gamma+1}{2 + (\gamma-1)Ma^2}\right]^{\frac{1}{\gamma-1}},$$

$$\frac{T}{T^*} = \frac{\gamma+1}{2 + (\gamma-1)Ma^2},$$

$$\frac{A}{A^*} = \frac{1}{Ma}\left[\frac{2 + (\gamma-1)Ma^2}{\gamma+1}\right]^{\frac{\gamma+1}{2(\gamma-1)}},$$

式中 p, ρ, T 和 Ma 分别是面积为 A 的截面上的压力、密度、温度和马赫数, 上标 * 号表示最小截面上的值. 管道中各物理量的分布可由进口条件惟一确定. 但是, 当管道出口反压大于以等熵流动理论算出的值时, 扩张段中可能会发生激波, 也有可能在最小截面处达不到音速, 因而在整个管道中都是亚音速的.

弹性塑性理论

弹性体(elastic body)　可变形体变形时, 每一点的应力张量与应变张量之间有相互确定的单值函数关系, 且当应力张量恢复为零张量时, 应变张量也恢复为零张量, 这种物体为弹性体. 当应力张量与应变张量之间为线性关系时, 称为线性弹性体. 在大多数情况下, 弹性体指的就是线性弹性体.

线性弹性理论(linear theory of elasticity)　弹性力学的重要分支. 在弹性材料的应力应变关系遵从胡克定律、应变与位移关系为柯西应变、平衡方程列在未变形物体的几何关系上、所有外力在变形过程中大小方向不改变的条件下, 弹性力学问题可看成线性弹性力学问题. 线性弹性力学问题成熟得较早, 而且在大多数情况下, 线性弹性问题是弹性力学问题很好的近似, 所以在不特别声明时, 一般说弹性理论, 大多指的是线性弹性理论.

非线性弹性理论(nonlinear theory of elasticity)　一种弹性理论. 是弹性力学中各物理量的关系引入非线性关系所建立的理论体系. 弹性力学中, 一般有三类基本关系, 即本构关系、应变位移关系和

平衡方程. 在本构关系中考虑非线性因素，称为物理非线性；在应变位移关系中引进非线性，或在平衡方程中考虑物体的几何变形，称为几何非线性. 有的弹性力学问题中兼有这两种非线性. 非线性弹性力学问题的求解比较复杂，通常采用近似方法或数值方法.

应力(stress)　弹性力学的基本概念之一. 指物体内在过一点微面积元两侧相互作用力的集度. 其大小和方向决定于点的位置和微面积元的方向. 令过某点的面积元为 $\Delta\boldsymbol{S}$，$\Delta\boldsymbol{S}$ 法线 $\boldsymbol{n}$ 所指的一侧作用于另一侧的合力为 $\Delta\boldsymbol{F}$，则这一点在 $\Delta\boldsymbol{S}$ 法方向上的应力为

$$\boldsymbol{\sigma} = \lim_{\Delta S\to 0}\frac{\Delta\boldsymbol{F}}{\Delta\boldsymbol{S}}.$$

$\sigma_n=\boldsymbol{\sigma}\cdot\boldsymbol{n}$ 即应力在法向量方向的投影，为正应力，应力在面积元内的投影为剪应力.

应力张量(stress tensor)　一种张量. 是描述物体内一点应力状态的张量. 它是一个三维二阶张量. 在直角坐标系中，它的分量记为 $\sigma_{ij}(i,j=1,2,3$，分别对应于 $x,y,z)$，其中 $\sigma_{ij}=\sigma_{ji}$，所以它是一个对称张量. 令过一点的微面积元的法向量为 $\boldsymbol{n}=(n_1,n_2,n_3)$，则在这个微面积元上的正应力为

$$\sigma_n = \sigma_{ij}n_in_j,$$

这里采用了约定求和. 若令

$$p_n^2 = \sum_{j=1}^{3}(\sigma_{ij}n_j)^2,$$

则在这个面积元上的切应力为

$$\tau_n = \sqrt{p_n^2-\sigma_n^2}.$$

σ_{ij} 中，$\sigma_{11},\sigma_{22},\sigma_{33}$ 分别称为 x,y,z 方向上的正应力，$\sigma_{ij}(i\neq j)$ 称为在法向沿 x,y,z 轴时微面积元上剪应力沿 j 轴的分量. 以上引进的应力张量称为柯西应力张量. 此外有时还用到第一和第二皮奥拉-基尔霍夫应力张量. 若它们的分量分别记为 τ_{ij}，$T_{ij}(i,j=1,2,3)$，它们与柯西应力张量之间的关系可表为

$$\tau_{ik} = g\sigma_{ij}\frac{\partial X_k}{\partial x_j},$$

$$T_{KL} = g\sigma_{ij}\frac{\partial X_K}{\partial x_i}\frac{\partial X_L}{\partial x_j},$$

其中 g 是变形梯度 F 的行列式，$x_i,X_i(i=1,2,3)$ 分别是变形后和变形前的坐标. 张量 T_{ij} 也是对称张量. 应力张量分量的单位是 $\mathrm{N/cm^2}$，或 $\mathrm{N/m^2}$(亦称 Pa).

剪应力(shear stress)　亦称切应力. 一种应力. 是物体内过一点的微面积元上应力在该面积元内的分量.

主应力(principal stress)　一种应力. 是物体内某一点以法向量为 $\boldsymbol{n}=(n_1,n_2,n_3)$ 的微面积元上剪应力为零时的正应力. 这时，$\boldsymbol{n}$ 的方向称为这一点的应力主方向. 一点在某一微面积元上的正应力，当面积的法向量 $\boldsymbol{n}$ 变化时，在应力主方向上取驻值. 对于一点的应力张量 $\sigma_{ij}(i,j=1,2,3)$，主应力一般有三个，它们满足方程

$$\| \sigma_{ij} - \sigma\delta_{ij} \| = 0,$$

这个三次方程的解即主应力 $\sigma_i(i=1,2,3)$. 对于给定点的应力张量，主应力是坐标变换下的不变量.

应力主方向(principal direction of stress)　连续介质力学概念. 指物体内的微面积元上剪应力为零时，该面积元的法方向. 令这个方向为 $\boldsymbol{n}^i=(n_1^i, n_2^i, n_3^i)$，对应的主应力为 $\sigma_i(i=1,2,3)$，这一点的应力张量为 $\sigma_{ij}(i,j=1,2,3)$，则有

$$(\sigma_{ij}-\sigma_l\delta_{ij})n_j^l = 0 \quad (l=1,2,3).$$

对一点，已知其应力状态，δ_{ij} 为克罗内克 δ，它的应力主方向有三个，而且是相互垂直的.

弹性力学平衡方程(equilibrium equation in elasticity)　联系应力张量和外力的平衡条件. 令 D 为变形体所占的区域，S 为它的表面，$\boldsymbol{n}=(n_1,n_2,n_3)$ 是表面外法线上的单位向量，$\boldsymbol{f}=(f_1,f_2,f_3)$ 为 D 内各点的外力，$\boldsymbol{t}=(t_1,t_2,t_3)$ 为 S 上的应力向量，σ_{ij} 为应力张量，则平衡方程为

$$\frac{\partial\sigma_{ij}}{\partial x_j}+f_i=0 \quad (\text{在 } D \text{ 内的每一点}),$$

$$\sigma_{ij}n_j+t_i=0 \quad (\text{在 } S \text{ 上每一点}).$$

在线性弹性力学中，考虑 D,S 和 x_i 为变形前的物体及其坐标. 在非线性弹性力学中 D,S,x_i 为变形后的物体及其坐标.

应变(strain)　变形体力学概念. 泛指应变张量的分量. 应变张量有格林应变张量(也称为拉格朗日应变张量)、阿尔曼西应变张量(也称为欧拉应变张量)，这二者是针对有限变形的. 还有柯西应变张量，是针对微小变形的. 一般称应变，多数情况是指柯西应变张量的分量，如正应变(或称线应变)、剪应变(或称切应变、角应变).

柯西应变(Cauchy strain)　一种应变. 是描述可变形体在一点的变形状态的一种应变张量. 最早由法国数学家、力学家柯西(Cauchy, A.-L.)在 1823 年引进. 在直角坐标系中，该张量分量与位移的关系为

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right),$$

这里 ε_{ij} 为应变张量分量，u_i 为位移分量. 它是一种线性关系定义的应变.

剪应变(shear strain)　一种应变. 指变形时，对过物体中一点且相互垂直的线素夹角的改变. 设在 xy 平面内，在 x,y 方向的位移分量分别为 u 和 v，略去高阶项变形，则两坐标向量的夹角变化为剪

应变，

$$r_{xy}=\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}=r_{yx}.$$

在数值上，它等于相应的应变张量分量 ε_{xy} 和 ε_{yx} 的二倍.

应变张量(strain tensor) 三维二阶张量. 是描述物体中一点的变形状态的张量. 在直角坐标系 $Oxyz$ 中，分量为 $\varepsilon_{ij}(i,j=1,2,3$ 分别对应于 x,y,z 三个方向)，当 $i\neq j$ 时，$\varepsilon_{ij}=\varepsilon_{ji}$，它是一个对称张量. 其中 $\varepsilon_{11},\varepsilon_{22},\varepsilon_{33}$ 分别对应于三个坐标轴方向上的伸长率，称为正应变. 当 $i\neq j$ 时，ε_{ij} 对应于 i 轴与 j 轴夹角变化值(即剪应变)的二分之一. 令物体上的位移场向量为 $\boldsymbol{u}(u_1,u_2,u_3)$，则应变张量分量与位移分量间的关系为

$$\varepsilon_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)\quad(i,j=1,2,3).$$

在有些情形下，上述线性关系不能满足要求，需要考虑非线性表达式. 若令 $X_i,x_i(i=1,2,3)$ 分别为变形前和变形后的坐标分量，$U_i,u_i(i=1,2,3)$ 分别为变形前和变形后位移，变形后的线素 $\mathrm{d}s$ 和变形前的线素 $\mathrm{d}S$ 的平方差为

$$\mathrm{d}s^2-\mathrm{d}S^2=2E_{ij}\mathrm{d}X_i\mathrm{d}X_j=2e_{ij}\mathrm{d}x_i\mathrm{d}x_j,$$

其中

$$E_{ij}=\frac{1}{2}\left(\frac{\partial U_i}{\partial X_j}+\frac{\partial U_j}{\partial X_i}+\frac{\partial U_l}{\partial X_i}\frac{\partial U_l}{\partial X_j}\right),$$

$$e_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}-\frac{\partial u_l}{\partial x_i}\frac{\partial u_l}{\partial x_j}\right),$$

E_{ij} 称为格林应变张量，e_{ij} 称为阿尔曼西应变张量.

主应变(principal strain) 一种应变. 指物体内某一点沿某方向 $\boldsymbol{n}=(n_1,n_2,n_3)$ 上的伸长率，且这个方向与它垂直的任何方向的剪切变形为零. 这个方向称为应变主方向. 一点沿方向 $\boldsymbol{n}$ 的伸长率，当 $\boldsymbol{n}$ 变化时，在应变主方向上取驻值. 对于一点的应变张量 $\varepsilon_{ij}(i,j=1,2,3)$，主应变一般有三个，它们满足方程

$$\|\varepsilon_{ij}-\sigma\delta_{ij}\|=0,$$

这个三次方程的解即为主应变 $\varepsilon_i(i=1,2,3)$. 对于给定的应变张量，一点的主应变是坐标变换下的不变量.

应变主方向(principal direction of strain) 物体内某一点对应于主应变的方向. 令这个方向 $\boldsymbol{n}^l=(n_1^l,n_2^l,n_3^l)$，对应的主应变为 $\varepsilon_l(l=1,2,3)$，这一点的应变张量为 ε_{ij}，则有

$$(\varepsilon_{ij}-\varepsilon_l\delta_{ij})n_j^l=0\quad(l=1,2,3).$$

对于一点已知应变状态，它的应变主方向有三个，而且是相互垂直的.

应变协调条件(compatibility condition of strain) 亦称应变协调方程. 线性弹性力学应变分量必须满足的一组微分关系式. 在直角坐标系 $Oxyz$ 中，应变协调方程是

$$\frac{\partial^2\varepsilon_{ij}}{\partial x_k\,\partial x_l}=\frac{\partial^2\varepsilon_{ik}}{\partial x_l\,\partial x_j}+\frac{\partial^2\varepsilon_{jl}}{\partial x_k\,\partial x_i}-\frac{\partial^2\varepsilon_{kl}}{\partial x_i\,\partial x_j},$$

式中 $i,j,k,l=1,2,3$ 分别对应于 x,y,z. 上面共得六个独立方程. 一般地，由六个应变分量 ε_{ij} 决定区域上光滑的位移场的必要条件是这些应变分量满足协调条件.

应变协调方程(compatibility equation of strain) 即"应变协调条件".

广义胡克定律(generalized Hooke law) 物体力学性质的本构方程. 它描述线性弹性体应力和应变的物理关系. 在绝热或等温条件下，无预应力的线性弹性体，若各向异性，则

$$\begin{bmatrix}\sigma_{11}\\\sigma_{22}\\\sigma_{33}\\\sigma_{23}\\\sigma_{31}\\\sigma_{12}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}&a_{14}&a_{15}&a_{16}\\a_{12}&a_{22}&a_{23}&a_{24}&a_{25}&a_{26}\\a_{13}&a_{23}&a_{33}&a_{34}&a_{35}&a_{36}\\a_{14}&a_{24}&a_{34}&a_{44}&a_{45}&a_{46}\\a_{15}&a_{25}&a_{35}&a_{45}&a_{55}&a_{56}\\a_{16}&a_{26}&a_{36}&a_{46}&a_{56}&a_{66}\end{bmatrix}\begin{bmatrix}\varepsilon_{11}\\\varepsilon_{22}\\\varepsilon_{33}\\\varepsilon_{23}\\\varepsilon_{31}\\\varepsilon_{12}\end{bmatrix},$$

式中 σ_{11} 和 ε_{11} 分别为以 x 为法线的截面上的正应力和相应的正应变；σ_{12} 和 ε_{12} 分别为该面上沿 y 方向的剪应力和相应的剪应变，余类推. 组成对称矩阵的 21 个系数 $a_{11},a_{12},\cdots$ 是相互独立的，对非均匀体 a_{ij} 为位置的函数，对均匀体 a_{ij} 为常数. 在这个方程中，若把系数矩阵换为

$$\begin{bmatrix}\lambda+2\mu&\lambda&\lambda&0&0&0\\\lambda&\lambda+2\mu&\lambda&0&0&0\\\lambda&\lambda&\lambda+2\mu&0&0&0\\0&0&0&2\mu&0&0\\0&0&0&0&2\mu&0\\0&0&0&0&0&2\mu\end{bmatrix},$$

就得到同样情况下各向同性弹性体的广义胡克定律. 展开得

$$\sigma_{11}=\lambda(\varepsilon_{11}+\varepsilon_{22}+\varepsilon_{33})+2\mu\varepsilon_{11},$$

$$\sigma_{23}=2\mu\varepsilon_{23},$$

余类推. 这种关系常表示为

$$\varepsilon_{11}=\frac{1}{E}[\sigma_{11}-\nu(\sigma_{22}+\sigma_{33})],$$

$$\varepsilon_{23}=\frac{1+\nu}{E}\sigma_{23},$$

余类推. 称 λ,μ 为拉梅系数，E 为杨氏模量，ν 为泊松比，

$$\lambda=\frac{\nu E}{(1+\nu)(1-2\nu)},\quad\mu=\frac{E}{2(1+\nu)}.$$

工程上记 $G=\mu$，称为剪切模量. 各向同性线性弹性体本构方程的独立系数的个数为 2.

弹性体变形和受力的关系最早由英国自然科学家胡克(Hooke, R.)于 1676 年以字谜形式提出，两

年后他解释为:弹性体的力与其伸长成比例,即一维情形的本构关系.19世纪前半叶,法国数学家、力学家柯西(Cauchy,A.-L.)和泊松(Poisson,S.-D.)曾对三维的关系进行了研究.1838年,英国数学家格林(Green,G.)用能量守恒定律推断线性的应力应变关系应有21个独立系数;1855年,汤姆森(Thomson,W.)用热力学定律证明了同样的结论,即上面关系式右边矩阵是对称的.这与实验结果一致.而各向同性关系可在上述各向异性关系的基础上用数学方法导出,即以上关系只需两个独立的系数.因三维的应力应变关系被看做是一维关系的推广,故称为广义胡克定律.服从这种线性关系的固体,通常称为胡克体,有时也简称该定律为胡克定律.

胡克定律(Hooke law)　线性弹性体所服从的基本规律.由英国自然科学家胡克(Hooke,R.)提出而得名.中国早在汉代郑玄《考工记》注中就有这一规律的记载.现今弹性力学中胡克定律表述为:令σ_{ij}, $\varepsilon_{ij}(i,j=1,2,3)$为应力张量和应变张量的分量,则它们之间存在关系

$$\sigma_{ij}=\sum_{k,l=1}^{3}c_{ijkl}\varepsilon_{kl}\quad(i,j=1,2,3),$$

这里$c_{ijkl}(i,j,k,l=1,2,3)$为弹性系数,由于$c_{ijkl}=c_{klij}$,弹性系数在最一般情况下有21个.对于各向同性的弹性体,胡克定律中只有两个常数是独立的,弹性关系可表示为

$$\varepsilon_{11}=\frac{1}{E}[\sigma_{11}-\nu(\sigma_{22}+\sigma_{33})],$$

$$\varepsilon_{22}=\frac{1}{E}[\sigma_{22}-\nu(\sigma_{11}+\sigma_{33})],$$

$$\varepsilon_{33}=\frac{1}{E}[\sigma_{33}-\nu(\sigma_{22}+\sigma_{11})],$$

$$\varepsilon_{12}=\frac{1}{2G}\sigma_{12},$$

$$\varepsilon_{23}=\frac{1}{2G}\sigma_{23},$$

$$\varepsilon_{31}=\frac{1}{2G}\sigma_{31},$$

其中E为杨氏模量,ν为泊松比.G为剪切模量,它与E,ν之间有关系$G=E/[2(1+\nu)]$.各向同性材料的胡克定律也可以写为

$$\sigma_{ij}=\lambda\theta\delta_{ij}+2G\varepsilon_{ij}\quad(i,j=1,2,3),$$

式中$\theta=\varepsilon_{11}+\varepsilon_{22}+\varepsilon_{33}$, δ_{ij}为克罗内克δ,λ和G称为拉梅常数,且λ,G与E,ν间有关系

$$\lambda=\frac{\nu E}{(1+\nu)(1-2\nu)},\quad G=\frac{E}{2(1+\nu)},$$

G有时也记为μ.

杨氏模量(Young modulus)　亦称纵向弹性模量.一个数.是线性弹性体在简单拉伸或压缩情况下正应力和纵向应变之间的比例系数.由英国学者杨(Young,T.)于1807年提出而得名.令σ为正应力,ε为纵向应变,E为杨氏模量,则有$\sigma=E\varepsilon$(参见"弹性体").由于伸长率是量纲一的量,E的量纲和应力的单位相同.

剪切模量(shear modulus)　一个数.是剪应力同剪应变的比例系数.令γ为剪应变,τ为剪应力,G为剪切模量,则有$\tau=G\gamma$(参见"胡克定律").

泊松比(Poisson ratio)　亦称横向变形系数.一个数.是线性弹性体在简单拉伸情况下,横向应变和纵向应变之间的比例系数,令σ为正应力,ε为纵向应变,ε'为横向应变,ν为泊松比,则有

$$\varepsilon'=\nu\varepsilon\quad和\quad\varepsilon=\frac{-\nu\sigma}{E},$$

E为杨氏模量.泊松比是量纲一的量,理论上讲各种材料的泊松比处于$-1<\nu<1/2$.当$\nu=1/2$时,材料不可压缩;当$\nu=-1$时,材料不可剪切变形.实际上通常材料ν总是正的.

泊松比是由法国数学家、力学家泊松(Poisson,S.-D.)于1829年提出的.

拉梅方程(Lamé equation)　亦称纳维方程.以位移表示的各向同性线性弹性力学的一组偏微分方程,即

$$G\nabla^2\boldsymbol{u}+(\lambda+G)\nabla(\nabla\cdot\boldsymbol{u})+\boldsymbol{f}-\rho\frac{\partial^2\boldsymbol{u}}{\partial\tau^2}=0,$$

式中$\boldsymbol{u}$,$\boldsymbol{f}$分别为位移和体力向量,λ,G为拉梅常数(参见"胡克定律"),ρ为密度,∇是梯度算子.这组方程于1829年由法国数学家、力学家泊松(Poisson,S.-D.)确立,他修正了1821年法国数学家、工程师纳维(Navier,C.(-L.-M.-H.))最早给出的含一个弹性系数的方程组,而法国数学家拉梅(Lamé,G.)于1833年给出了现在的形式.

弹性力学的初边值问题(initial and boundary value problem of elasticity)　弹性力学中动力学或静力学问题的待解方程组和定解条件.对于线性弹性力学情形,令弹性体占有空间区域D,它的边界为s,在D上每一点给定体力$\boldsymbol{f}=(f_1,f_2,f_3)$,则有:

平衡(或运动)方程为

$$\frac{\partial\sigma_{ij}}{\partial x_j}+f_i=0\left(\rho\frac{\partial^2u_i}{\partial t^2}\right),\tag{1}$$

应变位移关系为

$$\varepsilon_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right),\tag{2}$$

胡克定律为

$$\sigma_{ij}=c_{ijkl}\varepsilon_{kl},\tag{3}$$

其中u_i为位移向量分量,σ_{ij}为应力张量分量,ε_{ij}为应变张量分量,c_{ijkl}为弹性系数,ρ为密度.令在边界s_u上给定位移,s_σ上给定面力,且$s=s_u\cup s_\sigma$,则有

$$\boldsymbol{u}|_{s_u} = \boldsymbol{u}^* = (u_1^*, u_2^*, u_3^*), \quad (4)$$

$$\sigma_{ij} n_j|_{s_\sigma} = p_i^*, \quad (5)$$

这里 $\boldsymbol{n}=(n_1,n_2,n_3)$ 为边界 s_σ 上的外法向量. 在边条件(4),(5)下求解方程组(1),(2),(3)中的 $\boldsymbol{u}$, σ_{ij}, ε_{ij} 称为线性弹性力学的边值问题. 如果讨论的是弹性动力学问题,则在(1)中应考虑惯性项,并需补充适当的初条件. 线性弹性力学初边值问题这种提法,其解的存在性和惟一性在大部分情况下已得到证明. 当变形较大时,要遇到非线性弹性力学问题.

弹性力学惟一性定理(uniqueness theorem of elasticity) 线性弹性力学边值问题解的定理. 该定理断言:线性弹性体在小位移条件下,其平衡方程列在未变形的物体上,在给定的弹性体上和边界上的自平衡力系作用下,弹性体内的应变和应力场惟一确定. 如果边界的某区域上给定位移,则弹性体内的位移场也惟一确定.

弹性力学惟一性定理是由德国学者基尔霍夫(Kirchhoff, G. R.)于 1858 年确立的.

弹性力学通解(general solution of elasticity) 体力为零时的拉梅方程解的一般表示式. 最常用的有苏联力学家巴泊柯维奇(Папкович, П. Ф.)于 1932 年提出的巴泊柯维奇通解,即

$$\boldsymbol{u}=\boldsymbol{b}-\frac{1}{4(1-\nu)}\operatorname{grad}(b_0+\boldsymbol{r}\cdot\boldsymbol{b}),$$

这里 $\boldsymbol{u}$ 为弹性位移场,向量 $\boldsymbol{b}$ 的分量与函数 b_0 皆为调和函数,$\boldsymbol{r}=(x,y,z)$,ν 为泊松比. 此外还有苏联学者伽辽金(Галёркин, Б. Г.)提出的伽辽金通解,即

$$\boldsymbol{u}=\nabla^2\boldsymbol{a}-\frac{1}{2(1-\nu)}\operatorname{grad}\operatorname{div}\boldsymbol{a},$$

其中 $\boldsymbol{a}$ 的分量为双调和函数,∇^2 为拉普拉斯算子.

弹性体的变形能(strain energy of elastic body) 弹性力学概念. 弹性体变形后所储藏的能量. 令弹性体在空间占区域 D,变形能可表示为

$$U = \iiint_D W\mathrm{d}x\mathrm{d}y\mathrm{d}z,$$

这里 W 为变形能密度. 弹性体的变形能有时也称为应变能,它是一种势能.

变形能密度(density of strain energy) 亦称应变能密度. 弹性力学概念. 指弹性体变形后单位体积所包含的变形能. 对于弹性体内一点,令这点的应变张量为 ε_{ij},应力张量为 σ_{ij},则这点的变形能密度为

$$W = \int \sigma_{ij}\mathrm{d}\varepsilon_{ij},$$

积分从 $\varepsilon_{ij}=0$ 到 ε_{ij},沿着六维空间的路径进行,而且结果与路径无关. 对于线性弹性体,应力应变满足胡克定律,积分可直接得到,在各向同性条件下,表达式为

$$W=\frac{E}{2(1-\nu)}\Big[\Big(1+\frac{\nu}{1-2\nu}\Big)\theta^2+\frac{1}{8}(\varepsilon_{12}^2+\varepsilon_{23}^2+\varepsilon_{13}^2)-2(\varepsilon_{11}\varepsilon_{22}+\varepsilon_{22}\varepsilon_{33}+\varepsilon_{11}\varepsilon_{33})\Big],$$

式中 $\theta=\varepsilon_{11}+\varepsilon_{22}+\varepsilon_{33}$.

余变形能(anamorphosis energy) 亦称余应变能. 弹性力学概念. 指弹性体变形时应力改变量在相应应变上所作功的积分. 设弹性体占有三维空间的区域 D,则余变形能为

$$U^c = \iiint_D W^c\mathrm{d}x\mathrm{d}y\mathrm{d}z,$$

这里 W^c 为余变形能密度,定义为

$$W^c = \int \varepsilon_{ij}\mathrm{d}\sigma_{ij},$$

积分由 $\sigma_{ij}=0$ 到 σ_{ij},ε_{ij} 为应变. W^c 只是应力的函数,积分与路径无关. 对于线性弹性体,在各向同性时,有

$$W^c = \frac{1}{2E}[\sigma_{11}^2+\sigma_{22}^2+\sigma_{33}^2 - 2\nu(\sigma_{11}\sigma_{22}+\sigma_{22}\sigma_{33}+\sigma_{11}\sigma_{33}) + 2(1+\nu)(\sigma_{12}^2+\sigma_{23}^2+\sigma_{13}^2)].$$

在一般情形下,它和变形能密度 W 有关系

$$W + W^c = \sigma_{ij}\varepsilon_{ij}.$$

余应变能(complementary strain energy) 即"余变形能".

克拉伯龙定理(Clapeyrom theorem) 亦称应变能定理. 表征弹性体变形能与应变关系的定理. 克拉伯龙(Clapeyrom, B. P. E.)于 1858 年提出. 该定理一般叙述为:由自然的不受力状态开始缓慢加载的弹性体,经过近似平衡的过程达到平衡态,其应变能等于外力在过程中所做的功. 由此得到弹性体单位体积的应变能,即应变能表达式为

$$W = \frac{1}{2}[\sigma_{11}\varepsilon_{11}+\sigma_{22}\varepsilon_{22}+\sigma_{33}\varepsilon_{33} + 2(\sigma_{23}\varepsilon_{23}+\sigma_{31}\varepsilon_{31}+\sigma_{12}\varepsilon_{12})],$$

式中 σ_{ij} 和 ε_{ij} 分别为应力张量和应变张量. 该表达式适合于任意的线性弹性体.

柯恩不等式(Korn inequality) 线性弹性力学中的重要不等式. 该不等式是

$$\iiint_D W\mathrm{d}x\mathrm{d}y\mathrm{d}z > \alpha\iiint_D (\nabla\boldsymbol{u})^2\mathrm{d}x\mathrm{d}y\mathrm{d}z,$$

式中 W 为变形能密度,D 为弹性体所占区域,α 为只与 D 有关的正实数,$\nabla\boldsymbol{u}$ 是位移向量 $\boldsymbol{u}$ 的梯度,即有

$$(\nabla\boldsymbol{u})^2 = \sum_{i,j=1}^{3}\frac{\partial u_i}{\partial x_j}\frac{\partial u_i}{\partial x_j}.$$

柯恩不等式在证明线性弹性力学解的存在性和近似解的误差估计中有许多应用. 柯恩不等式最早由柯恩(Korn, A.)于 1906 年证明,其后弗里德里希斯

(Friedrichs, K. O.)于 1947 年、米赫林(Михлин, С. Г.)于 1951 年和弗凯拉(Fichera, G.)分别给了不同的证明和讨论.

卡斯提利亚诺定理(Castigliano theorem)　由弹性体位移确定力以及由力确定位移的定理:平衡弹性体的应变能对某广义位移的变化率,等于所在点沿该广义位移方向的广义力;平衡弹性体的余应变能对某广义力的变化率,等于所在点沿该广义力方向的广义位移. 广义位移含转角,广义力含力偶矩,均由依右手法则规定的向量表示,应变能和余应变能分别以广义位移和广义力表示. 该定理适用于线性和非线性弹性力学问题. 对线性问题,余应变能就等于应变能,计算时只需将应变能表示为广义力的函数. 在杆件结构力学中,该定理应用较广,成为位移法和力法的理论基础. 通常称前后两部分依次为卡氏第一定理和第二定理,由卡斯提利亚诺(Castigliano, A.)于 1873 年提出. 对于第二定理,当时针对线性问题未区分应变能和余应变能. 关于余应变能对力求导而确定位移的内容,克罗蒂(Crotti, F.)于 1878 年、恩盖塞(Engesser, F.)于 1889 年分别提出,故卡氏第二定理也称为克罗蒂-恩盖塞定理.

布森内斯克解(Boussinesq solution)　均匀各向同性线性弹性力学问题的一个重要解. 弹性体为半无限体,表面为平面,在原点作用法向集中力,其解因由法国力学家、理论物理学家布森内斯克(Boussinesq, J. V.)提出而得名. 令在点(x,y,z)的位移为(u,v,w),则有

$$u=\frac{p}{4\pi G}\left[\frac{xz}{r^3}-(1-2\nu)\frac{x}{r(r+z)}\right],$$

$$v=\frac{p}{4\pi G}\left[\frac{yz}{r^3}-(1-2\nu)\frac{y}{r(r+z)}\right],$$

$$w=\frac{p}{4\pi G}\left[\frac{z^3}{r^3}+2(1-\nu)\frac{1}{r}\right],$$

式中 p 为法向荷载,沿 z 的正方向,G 为剪切模量,ν 为泊松比,$r=\sqrt{x^2+y^2+z^2}$,弹性体处于 xy 平面的下方,即 $z<0$ 一方.

弹性力学的开尔文解(Kelvin solution of elasticity)　在一点受集中力作用的全空间均匀各向同性线性弹性力学问题的一个重要解. 由英国学者开尔文(Kelvin, L.)于 1848 年求得解位移表达式是

$$u_{ij}=\frac{1}{16G\pi(1-\nu)}\left[\frac{3-4\nu}{\rho}\delta_{ij}+\frac{1}{\rho^3}(x_i-\xi_i)(x_j-\xi_j)\right],$$

其中 G 为剪切模量,ν 为泊松比,(ξ_1,ξ_2,ξ_3)为集中力作用点,

$$\rho=[(x_1-\xi_1)^2+(x_2-\xi_2)^2+(x_3-\xi_3)^2]^{\frac{1}{2}},$$

u_{ij}表示在 x_i 轴方向作用单位力的 x_j 轴方向的位移分量.

额舍耳比解(Eshelby solution)　弹性力学问题的一个特殊解. 当均匀各向同性线性弹性介质充满整个空间,而且具有初均匀应变 e_{jk}^T 时,弹性体内有一椭球形空洞,位移场为

$$u_i^T(x,y,z)=\frac{e_{jk}^T}{8\pi(1-\nu)}\iiint\frac{g_{ijk}(l)}{r^2}\mathrm{d}\xi\mathrm{d}\eta\mathrm{d}\zeta,$$

其中 ν 为泊松比,$g_{ijk}(l)=(1-2\nu)(\delta_{ij}l_k+\delta_{ik}l_j-\delta_{jk}l_i)+3l_il_jl_k$,$r$ 为由(ξ,η,ζ)到(x,y,z)的距离,而 l_1,l_2,l_3 为 r 的三个方向,积分在空洞外全空间进行. 该解在弹性体有微孔时估计材料的损伤特性中有重要的应用.

额舍耳比解是由英国学者额舍耳比(Eshelby, J. D.)于 1957 年得到的.

弹性力学的变分原理(variational principles of elasticity)　弹性力学的重要原理. 是把弹性力学的初边值问题的部分或全部的方程和初边条件等价于某一泛函的极值或驻值条件的原理. 在线性弹性力学静力学问题中经常用到的是以下一些变分原理:

1. 最小势能原理. 弹性体上的总势能写为以位移 u_i 为自变函数的泛函:

$$\pi(u_i)=\iiint_D W\mathrm{d}x\mathrm{d}y\mathrm{d}z-\iiint_D f_iu_i\mathrm{d}x\mathrm{d}y\mathrm{d}z-\iint_{s_\sigma}\bar{p}_iu_i\mathrm{d}s,$$

式中 D 为弹性体所占有的区域,s_u 为给定位移的边界,s_σ 为给定外力 $\bar{p}_i$ 的边界,W 为变形能密度,$\pi(u_i)$取极小值的充分必要条件是位移满足拉梅方程和 s_σ 上的应力边条件. u_i 的可变函数类要求在 s_u 上满足位移边条件,f_i 为体力.

2. 最小余能原理. 给出以应力张量 σ_{ij}为自变函数的泛函,即弹性总余能

$$\pi^c(\sigma_{ij})=\iiint_D W^c\mathrm{d}x\mathrm{d}y\mathrm{d}z-\iint_{s_u}p_i\bar{u}_i\mathrm{d}s,$$

式中 W^c 为余变形能密度,s_u 为 D 上给定位移 $\bar{u}_i$ 的边界. 当 σ_{ij}满足 s_σ 上应力边条件时,它使 π^c 取最小值的充分必要条件是 σ_{ij}满足位移连续条件和 s_u 上的位移边条件.

3. 赫林格-瑞斯纳变分原理,又称二类变量广义变分原理. 由赫林格(Hellinger, E.)于 1914 年和瑞斯纳(Reissner, E.)于 1950 年提出而得名. 该原理给出以 u_i 和 σ_{ij}为自变函数的泛函,即

$$\pi_2=\iiint_D\left[\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)\sigma_{ij}-W^c(\sigma_{ij})\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z-\iint_{s_u}(u_i-\bar{u}_i)p_i\mathrm{d}s-\iint_{s_\sigma}\bar{p}_iu_i\mathrm{d}s.$$

此泛函取驻值的充分必要条件是 σ_{ij},u_i 满足全部边条件和平衡方程以及协调方程.

4. 胡海昌-鹫津久一郎变分原理,又称三类变量

广义变分原理.由中国力学家胡海昌于1954年和日本学者鹫津久一郎于1955年各自独立提出而得名，该原理给出以 $u_i,\sigma_{ij},\varepsilon_{ij}$ 为自变函数的泛函，即

$$\pi_3=\iiint_D[W(\varepsilon_{ij})-f_iu_i]\mathrm{d}x\mathrm{d}y\mathrm{d}z-\iiint_D\left[\varepsilon_{ij}-\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)\right]\sigma_{ij}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z-\iint_{s_u}(u_i-\overline{u}_i)p_i\mathrm{d}s-\iint_{s_\sigma}\overline{p}_iu_i\mathrm{d}s.$$

它取驻值的充分必要条件是 $u_i,\varepsilon_{ij},\sigma_{ij}$ 满足全部弹性力学问题的方程和边条件.

后两个变分原理也称为广义变分原理.弹性力学变分原理可以同样用于梁、板、壳以及各种弹性力学的专门问题.这些原理被广泛用于推导各种特殊问题所满足的方程和边条件，以及各种近似计算中.

弹性力学最小势能原理(principle of minimum potential energy of elasticity)　见“弹性力学的变分原理”.

最小余能原理(principle of minimum complementary energy)　见“弹性力学的变分原理”.

赫林格-瑞斯纳变分原理(Hellinger-Reissner variational principle)　见“弹性力学的变分原理”.

胡海昌-鹫津久一郎变分原理(Hu-Washizu variational principle)　见“弹性力学的变分原理”.

弹性力学广义变分原理(generalized variational principles of elasticity)　见“弹性力学的变分原理”.

功的互等定理(reciprocal theorem of work)　亦称互等功定理.关于线性弹性力学问题两组不同解的关系的定理.该定理断言：作用在弹性体上的两组广义力 F_a,F_b 对应的位移解分别为 u_a,u_b，则 F_a 在 u_b 上做的功等于 F_b 在 u_a 上做的功，即有 $\langle F_a,u_b\rangle=\langle F_b,u_a\rangle$，$\langle,\rangle$ 表示内积.功的互易定理对于由许多弹性体组成的复杂系统也是成立的.

功的互等定理是由意大利学者贝蒂(Betti，E.)于1872年和英国学者瑞利(Rayleigh，L.)于1873年分别提出.

坎托罗维奇法(Kantorovich method)　求解泛函极值问题(即变分问题)的一种直接方法.是苏联数学家、数理经济学家坎托罗维奇(Канторович，Л. В.)于1950年提出的.直接法用于力学中，是把泛函极值问题看做无限多个变量的函数极值问题，即有限个变量极值问题的极限情况，利用变分原理求解.实用上，按有限个变量处理，就是一种近似法.坎托罗维奇是把问题转化为 m 个单变量函数的选取，m 愈大精度愈高.它用于依赖于多变量函数的泛函.不失一般性，设泛函

$$\Phi[Z(x,y)]=\iint_{(x,y)}F[x,y,Z,Z_x,Z_y]\mathrm{d}x\mathrm{d}y.$$

该方法取

$$Z_m=\sum_{k=1}^{m}\alpha_k(x)w_k(x,y),$$

使满足 $Z(x,y)$ 的边条件，式中坐标函数 $w_k(x,y)$ 已选定，$\alpha_k(x)$ 待定，是多个自变量中一个变量的函数.代入泛函，化为

$$\Phi[Z_m]=\int_{x_0}^{x_1}\varphi[x,\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x),\alpha_1'(x),\alpha_2'(x),\cdots,\alpha_m'(x)]\mathrm{d}x,$$

把问题变为选取 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$，使 Φ 达到极值.在函数 $w_k(x,y)$ 的选取上，该方法类似里茨法，只是把 Z_m 表达式中的常数 α_k 换为函数.从而减少了选取坐标函数带来的影响，但增加了计算量.为了进一步消除坐标函数的影响，后人对该方法又有新的推广和改进.

弹性力学平面应变问题(plane strain problem of elasticity)　弹性力学的一类特殊问题.在弹性力学的初边值问题中，如果各方程和初边条件中的已知和未知量都不含 z 坐标，且位移与外力的 z 方向分量为零，应力、应变分量中，

$$\varepsilon_{33}=\varepsilon_{13}=\varepsilon_{23}=\sigma_{13}=\sigma_{23}=0,$$

这样的弹性力学问题称为平面应变问题.

在平面应变问题中，如果没有体力，引进应力函数 $\Phi(x,y)$，使

$$\sigma_{11}=\frac{\partial^2\Phi}{\partial y^2},\quad\sigma_{22}=\frac{\partial^2\Phi}{\partial x^2},\quad\sigma_{12}=\frac{\partial^2\Phi}{\partial x\,\partial y},$$

则平衡方程恒满足.在各向同性的条件下，将应力代入胡克定律，再把得到的应变通过应力函数的表达式，代入相应的应变协调方程，使得

$$\frac{\partial^4\Phi}{\partial x^4}+2\frac{\partial^4\Phi}{\partial x^2\,\partial y^2}+\frac{\partial^4\Phi}{\partial y^4}=0.$$

这说明，在没有体力时，各向同性线性弹性平面应变问题的应力函数是双调和函数.平面应变问题适用于无限长的柱体，及只受垂直于母线的外载的情形.

弹性力学平面应力问题(plane stress problem of elasticity)　弹性力学的一类特殊问题.在弹性力学的初边值问题中，所有方程和初边条件中的已知、未知量都不依赖于 z 坐标，沿 z 方向的外力为零，且 $\sigma_{13}=\sigma_{23}=\sigma_{33}=0$，这样的弹性力学问题称为弹性力学平面应力问题.在没有体力的条件下，与平面应变问题类似，应力函数 $\Phi(x,y)$ 满足双调和方程.在不假定沿 z 方向的位移和 $\sigma_{3i}(i=1,2,3)$ 为零的条件下，平面应力问题的解是弹性力学问题解的一种近似.因为它不满足全部协调条件，只有在弹性体沿 z 方向尺度很小时，近似程度才较好.

弹性力学平面应变问题和弹性力学平面应力问

题统称为弹性力学平面问题.

弹性力学平面问题(plane problem of elasticity)　见“弹性力学平面应力问题”.

应力函数(stress function)　亦称艾里应力函数.微分方程的一种解.是定义在弹性体所占区域上的函数.其应力分量是它的某些给定的微分表达式,能使弹性力学无体力的平衡方程恒满足.最早由英国学者艾里(Airy,G. B.)于1863年将其引入平面弹性问题.令 $\Phi(x,y)$ 为应力函数,并令

$$\sigma_{xx}=\frac{\partial^2\Phi}{\partial y^2},\quad \sigma_{xy}=-\frac{\partial^2\Phi}{\partial x\,\partial y},\quad \sigma_{yy}=\frac{\partial^2\Phi}{\partial x^2},$$

则应力恒满足平衡方程

$$\frac{\partial\sigma_{xx}}{\partial x}+\frac{\partial\sigma_{xy}}{\partial y}=0,$$

$$\frac{\partial\sigma_{xy}}{\partial x}+\frac{\partial\sigma_{yy}}{\partial y}=0.$$

后来,英国物理学家麦克斯韦(Maxwell, J. C.)和莫瑞拉(Morera,G.)分别于1868年和1892年提出空间弹性力学问题其他形式的应力函数.

艾里应力函数(Airy stress function)　即“应力函数”.

弹性力学复变函数方法(complex variable function method of elasticity)　基于复变函数理论的求解弹性力学问题的方法.主要是针对求解弹性力学平面问题的.最早由俄国学者科拉索夫(Колосов, Г. В.)于1914年提出,后经苏联学者穆斯赫利什维里(Мусхелишвили, Н. И.)系统发展而成熟.方法是将双调和方程写为复数形式

$$\frac{\partial^4\Phi}{\partial z^2\,\partial\bar{z}^2}=0,$$

Φ 为应力函数,$z=x+\mathrm{i}y$,$\bar{z}$ 为 z 的共轭复数.它的通解为

$$\Phi=\mathrm{Re}[\bar{z}\psi(z)+\chi(z)],$$

$\psi(z)$,$\chi(z)$为任意解析函数.于是,将问题归结为在给定边条件下求 $\psi(z)$和 $\chi(z)$.复变函数方法被广泛应用于求解平面应力集中问题、断裂力学问题,后来还被推广应用于特殊的弹性力学空间问题和板壳问题.

弹性柱体扭转(torsion of elastic cylinder)　一类弹性力学问题.指弹性柱体在端头力偶作用下的扭转问题.最早由法国力学家、几何学家圣韦南(Saint-Venant, A. J. C. B. de)于1855年研究过,其后德国学者普朗托(Prandtl, L.)于1903年和俄国学者于1913年用不同方法分别予以解决.弹性柱体扭转问题的解决,一方面基于圣韦南原理,即在端头放松边条件;另一方面基于先假设问题的一部分未知量为已知,然后求未知的部分,即所谓半逆解法.设柱截面为 xy 平面上的区域 D,母线沿 z 轴.令

$$\sigma_x=\sigma_y=\sigma_{xy}=\sigma_z=0,$$

且

$$\sigma_{xz}=\frac{\partial\Phi}{\partial y},\quad \sigma_{yz}=\frac{-\partial\Phi}{\partial x},$$

Φ 为扭转应力函数,将这些量代入弹性关系,并求出应变代入协调条件,则有

$$\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}=-2G\theta,$$

式中 G 为拉梅系数之一,亦称剪切模量,θ 为柱单位长度的扭角.Φ 在 D 的边界 s 上满足边条件 $\Phi|_s=0$.求解这个边值问题,可得未知应力 σ_{xz},σ_{yz},进而可以通过积分求得端头的扭矩 M_z.以上这种提法的扭转问题称为自由扭转问题.一般情况下,除圆截面外,扭转后横截面要发生翘曲,即不再保持为平面.如果在一端加以约束,使端面保持平面,这种问题称为约束扭转问题.

平截面假定(assumption of plane cross-section)　对于细长可变形柱体变形的一个基本假定.其内容是:变形前垂直于轴线的平面,即横截面上的质点,在变形过程中它们彼此距离不变,始终处于一个平面上,而且这个平面与变形后的轴线保持垂直.它适用于杆的拉伸、弯曲和圆截面杆的扭转.平截面假定,可将三维弹性体或塑性体等可变形细长柱体化归为一维问题处理,从而使问题大大简化,而且近似程度很好.

圣韦南原理(Saint-Venant principle)　在弹性力学中对于力的作用的局部效应现象的总结.该原理断言:作用在弹性体上一小块面积(或体积)上的力系,如果用另外一组与之静力等效的(即合力和合力矩相等)力系来代替,则所产生的应力改变,随与该小块面积的距离的增加而迅速减小.在求解实际弹性力学问题时,如果精确的应力边界条件难于提出,经常在小面积上用等效的已知应力分布来代替,得到的解只在小面积邻近有差异.圣韦南原理是由法国力学家、几何学家圣韦南(Saint-Venant, A. J. C. B. de)于1855年提出的.

薄膜比拟(membrane analogy)　求扭转应力的一种实验方法.是利用薄膜的函数表达式和应力函数的相似性求扭转应力的方法.对于弹性直杆纯扭转,在离施力端相当远时,只受垂直于杆轴的切应力,且与横截面位置无关.取坐标系 $Oxyz$,令 z 轴与杆平行,则6个应力分量中只有 τ_{zx},τ_{zy}不恒为零且与 z 无关.取 $\varphi(x,y)$为应力函数,

$$\tau_{zx}=\frac{\partial\varphi}{\partial y},\quad \tau_{zy}=-\frac{\partial\varphi}{\partial x},$$

平衡方程自然满足.由协调关系得

$$\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}=-2\mu\alpha,$$

式中 μ,α 分别为杆的剪切模量、单位长度扭转角,在边界可使 $\varphi(x,y)=0$. 横截面上任一点的切应力 τ 必沿 $\varphi(x,y)=$常数在相应位置的切向,

$$|\tau| = \left|\frac{\partial\varphi}{\partial n}\right|,$$

n 为该曲线的法线方向. 扭矩

$$M = 2\iint\varphi\mathrm{d}x\mathrm{d}y.$$

若截面为单连通,可作形状相同的空洞,沿其边张薄膜,施内压 q,使挠度微小,在洞平面上取坐标系 Oxy,则薄膜可表示为 $z=f(x,y)$,满足

$$\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}=-\frac{q}{T},$$

式中 T 为膜之张力. 可见 f 与 φ 相似. 故杆的切应力方向就对应薄膜等高线的切向,且

$$|\tau| = \frac{2\mu\alpha T}{q}\left|\frac{\partial f}{\partial n}\right|,$$

而扭矩

$$M = \frac{4\mu\alpha T}{q}\iint f\mathrm{d}x\mathrm{d}y.$$

在测得

$$\frac{q}{T},\quad \iint f\mathrm{d}x\mathrm{d}y,\quad \frac{\partial f}{\partial n}$$

后,就可由 M 而确定 α 和 τ 之值. 若再于半径为 r 的圆洞上张同样的薄膜,施同样内压 q,则

$$\frac{q}{T}=\frac{2\sin\beta}{r},$$

β 为膜在圆洞边的倾角,较易测得. 若扭杆有柱形空洞,可在实验模型上和洞对应地张起刚性轻薄片,使其保持平行于模型平面并与薄膜连结,再施内压,其余均不变. 薄膜比拟法由德国学者普朗托(Prandtl, L.)于 1903 年提出,对截面形状复杂的扭杆很有用处,很多情况能直观表示"应力丘"的形状,对切应力分布情况给出某些结论,甚至可导出这些应力的近似计算公式.

弹性薄板(elastic thin plate)　一类弹性体. 是厚度比另外两个尺度小很多的弹性体,或介于两平行平面之间的弹性体. 两平面间的距离 h 称为厚度,和两平面等距的平面称为板的中面,中面上的区域为 $\mathscr{D}$,h 比起 $\mathscr{D}$ 的最小直径小一个数量级时,就是薄板. 弹性力学平面应力问题的对象,在外形上也是薄板,但通常指的薄板问题,是指它的外载荷垂直于中面,它的变形主要是弯曲变形. 描述它的变形通常用中面上点的法向上的位移表示,这个位移称为挠度. 令中面所在的平面为 xy 平面,则挠度 $w(x,y)$ 满足方程

$$D\Delta^2 w = q(x,y)\quad(\mathscr{D}\text{ 上}),$$

式中 $q(x,y)$为分布荷载,$D=Eh^3/12(1-\nu^2)$称为板的抗弯刚度,E 为杨氏模量,ν 为泊松比,Δ^2 为双调和算子. 板的挠度方程在适当的边条件下可以求解,通常是在 $\mathscr{D}$ 的边界上给定挠度 w、绕边界转角 $\frac{\partial w}{\partial n}$、弯矩 M_n 和剪力 Q_n 四个量中的两个,M_n 和 Q_n 可以表示为 w 的微分表达式. 如果在板的边界上作用有板面内的外力,板内的应力可以由弹性力学平面应力问题求解得到,设为 $\sigma_x,\sigma_y,\tau_{xy}$,则这时板的挠度方程为

$$D^2 w=q(x,y)+h\left(\sigma_x\frac{\partial^2 w}{\partial x^2}+2\sigma_{xy}\frac{\partial^2 w}{\partial x\,\partial y}+\sigma_y\frac{\partial^2 w}{\partial y^2}\right).$$

板的弯曲问题,在工程上有许多应用,求解它比较困难,经常采用近似方法和数值方法. 弹性薄板问题,除了求解平衡问题外,还有振动、稳定性等其他问题.

直法线假定(assumption of straight line)　对于板和壳变形的一个基本假定. 最早由德国学者基尔霍夫(Kirchhoff, G. R.)对于薄板问题提出,1888 年,英国弹性力学家拉弗(Love, A. E. H.)推广到薄壳问题,所以又称为基尔霍夫-拉弗假定. 其内容是:变形前处于板或壳中面法线上的质点,在变形过程中始终处于同一直线上,这一直线保持垂直于变形后的中面,而且这些质点彼此之间的距离不变.

直法线假定可以将作为三维变形体问题的板壳化归为二维问题,从而使问题大大简化,而且当厚度较小时精度很好.

弹性薄壳理论(theory of elastic thin shell)　弹性力学的一个分支. 弹性薄壳是指与一块曲面段的距离小于 $h/2$ 的点组成的弹性体,曲面段称为薄壳的中面,h 称为壳厚,一般设壳厚比起曲面段的其他尺寸,如曲率半径、跨度等小一个数量级. 弹性薄壳理论是利用直法线假定,将这个三维弹性体的问题化归为定义在曲面段上的二维问题来求解. 它是由英国弹性力学家拉弗(Love, A. E. H.)在 20 世纪初完成的. 后来不同学者采用不同的具体路径来实现,在略去什么样的小量时略有区别. 基本方法是:在曲线坐标系中列出三维弹性体的方程组,其中坐标轴 z 沿曲面的法方向. 假设法向应力 $\sigma_{zz}=0$,变形服从直法线假设,即 $\varepsilon_{zz}=\varepsilon_{xz}=\varepsilon_{yz}=0$,于是整个问题可化归为在曲面段上求 15 个未知函数,即位移向量的三个分量、六个变形分量和六个内力分量. 其中六个变形分量的三个表征曲面的拉伸与剪切称为膜应变,另外三个称为弯扭变形分量;相应地六个内力分量的三个称为薄膜力,另外三个称为弯扭矩. 由于所得到的方程组比较复杂,只有在比较特殊的条件下才能得到准确解,如柱壳的轴对称变形等. 许多重要的工程结构都是由薄壳构成,所以薄壳理论有时作为结构力学的分支. 求解薄壳理论方程一般采用近似方法,有时也作进一步简化,如略去其中的弯扭

矩，这种近似称为薄膜理论. 当中面为平面时，可以把板的弯曲问题和平面应力问题看做薄壳问题的特殊情形. 如果构成薄壳的材料不是弹性材料，而是塑性、粘性或其他非线性材料，则薄壳问题变得更为复杂，成为物理非线性薄壳问题. 如果薄壳的中面的位移较大，以至需在变形分量中考虑非线性项，就形成几何非线性问题，或称薄壳的大挠度问题. 有时还需要讨论薄壳的动力问题，如振动和波的问题，这时，在问题的提法上要计入惯性力影响.

薄膜理论(theory of membrane)　亦称薄壳的无矩理论. 薄壳理论的一种简化理论，是在薄壳平衡方程中略去弯扭矩. 在薄壳十分薄的情况下，对实际问题有比较好的近似. 除了在曲率间断、外载间断、加肋条和固定边界的邻近外，比较薄的壳，薄膜力引起的应力占主要地位. 薄膜力的三个平衡方程可以化归为一个二阶偏微分方程，方程的类型随曲面上的点为椭圆点、双曲点、抛物点，相应为椭圆型方程、双曲型方程和抛物型方程. 求解这个方程一般采用数值方法，对于椭圆型情形，有时可以用广义解析函数工具求解.

卡门方程(Kármán equation)　亦称薄板大挠度方程组. 描述薄板挠度的方程. 由美籍匈牙利力学家卡门(Kármán，T. von)于 1910 年导出. 用于随着挠度增大，板中面内的应力不可忽略的情形. 它们是如下的一对偏微分方程：

$$D\Delta^2 w=q(x,y)+h\left(\frac{\partial^2\varphi}{\partial y^2}\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2\varphi}{\partial x^2}\frac{\partial^2 w}{\partial y^2}-2\frac{\partial^2\varphi}{\partial x\,\partial y}\frac{\partial^2 w}{\partial x\,\partial y}\right),$$

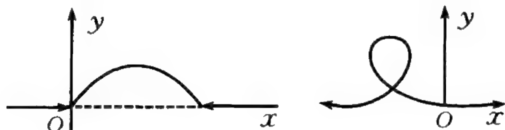
$$\Delta^2\varphi=E\left[\left(\frac{\partial^2 w}{\partial x\,\partial y}\right)^2-\frac{\partial^2 w}{\partial x^2}\frac{\partial^2 w}{\partial y^2}\right],$$

式中 $q(x,y)$ 为垂直作用于中面上的分布载荷，w 为挠度，h 为板厚，

$$D=\frac{Eh^3}{12(1-\nu^2)},$$

E 为杨氏模量，ν 为泊松比，Δ^2 为双调和算子，φ 为应力函数. 补充适当的边条件，可以求解这组方程. 通常多用近似方法求解.

欧拉弹性线(Euler elastic line)　弹性直杆受压的大变形. 弹性等截面细直杆，当轴向压力超过欧



拉临界载荷时的弯曲形式. 由瑞士数学家欧拉(Euler，L.)于 1744 年得到. 问题归结于积分 $EJk+py=0$，其中 EJ 为弯曲刚度，k 为曲率，p 为轴压，y 为挠度. 在两端相应的边条件下，其解的杆曲线为

$$x=\frac{1}{k}[2(E-E(\varphi))-(F-F(\varphi))],$$

$$y=\frac{2m}{k}\cos\varphi,$$

其中

$$E(\varphi)=\int_0^{\varphi}\sqrt{1-m^2\sin^2\varphi}\,\mathrm{d}\varphi,$$

$$F(\varphi)=\int_0^{\varphi}\frac{\mathrm{d}\varphi}{\sqrt{1-m^2\sin^2\varphi}},$$

$$E=E\left(\frac{\pi}{2}\right),\quad F=F\left(\frac{\pi}{2}\right),$$

$$k^2=\frac{p}{EJ},\quad m=\sin\frac{\theta_0}{2},$$

θ_0 为曲线在杆端与 x 轴的夹角.

弹性系统的稳定性理论(stability theory of elastic system)　弹性力学的分支学科. 是关于弹性体系或弹性结构的平衡或运动状态是否稳定的理论. 其内容包括：弹性系统平衡或运动稳定性的定义、判别准则，以及在各种具体情形下通过分析或数值方法进行判别的技巧. 弹性系统的平衡或运动状态给以微小扰动时，扰动后的运动状态与原来状态的偏差也是微小的，则这种状态称为是稳定的，否则称为是不稳定的. 由于弹性系统的平衡和运动可以用有限或无限自由度动力系统来描述，所以动力系统稳定性理论在这里也是适用的，但针对弹性系统的特点，还有其独特的内容，如稳定性准则中的静力准则、能量准则、关于稳定性的初缺陷理论等. 严格说来，弹性系统稳定性理论是和非线性弹性力学相联系的. 线性弹性系统的解自然是稳定的. 只有在弹性力学的诸方程中引入非线性因素才会产生不稳定的情形.

次弹性(hypoelasticity)　亦称低弹性. 一种弹性体. 指应力张量的某种导数同变形速率张量之间有线性关系的弹性体. 即 $\dot{\boldsymbol{\sigma}}=H(\boldsymbol{\sigma})\boldsymbol{D}$，其中 $\boldsymbol{D}$ 为变形速率张量，$\dot{\boldsymbol{\sigma}}$ 为应力张量 $\boldsymbol{\sigma}$ 的导数，该导数的定义有多种方式，又称为本构导数，它考虑观察者随物质团一起平移和转动. H 为次弹性物质的响应函数. 在一定条件下，普通的线性弹性本构、塑性中的增量理论可以看做它的特殊情形，所以次弹性本构模型是一种十分广的模型.

低弹性(low elasticity)　即“次弹性”.

超弹性(superelasticity)　一种弹性体. 具有变形能密度函数的变形体. 它的本构关系可表示为

$$\sigma_{ij}=\frac{\partial w}{\partial\varepsilon_{ij}},$$

这里 w 是变形能密度函数，它是应变张量的函数. σ_{ij}，ε_{ij}分别为应力、应变张量. 在一般情形下，由它确定的应力、应变关系不再是线性的. 当 w 是线性弹性体的应变能密度时，应力、应变关系就是胡克定

律.

粘弹性理论(theory of viscoelasticity) 连续介质力学的一个分支.是在本构方程中考虑弹性和粘性性质的理论.通常,其应力、应变关系可写为 $\varepsilon = f(\sigma,\dot{\sigma})$,其中 $\varepsilon,\dot{\sigma}$ 分别是应变与应力速率,σ 为应力.如果 f 为线性函数,则称为线性粘弹性理论.通常的固体材料,如混凝土、高分子材料、高温下的金属等,都会表现出在应力不变时的应变随时间变化的性质,因而都可以进一步近似地看做粘弹性材料.线性粘弹性本构关系的一般形式可写为

$$\sigma + a\dot{\sigma} = E(\varepsilon + b\dot{\varepsilon}),$$

式中 a,b,E 为常数.此单向的应力、应变关系也可以推广到三维情形.

温度应力(thermal stress) 亦称为变温应力或热应力.涉及温度问题的应力.是温度变化在弹性体内引起的应力.温度因时间或因位置的变化必使弹性体发生胀缩变形,边界的约束以及各局部变形的不均匀都可导致体内各部分相互作用,即产生应力.不少情况下这种应力的影响比较小,可忽略不计.但工程上,如土工建筑、化工设备、高速飞行器、各种热交换器等,温度变化很显著,易引发不可避免的力学问题,温度应力就成为分析研究的重点.考虑温度应力的弹性力学称为热弹性力学.

热弹性力学(thermoelasticity) 弹性力学的一个分支.是温度变化的效应不能忽略的弹性力学.一般地,弹性体会因温度随时间或位置变化而产生温度应力,只有这种应力影响不大时才可忽略.热弹性力学的问题一般分确定温度场和由温度场的变化来确定应力场两部分.后者主要解决温度随时间变化引起的应力问题,也可分两类:定常问题,温度应力不随时间而变;非定常问题,温度应力随时间而变,如热冲击、热弹性振动等.热弹性力学基本方程仍为平衡(动力)方程、几何方程和物理方程.前两种方程和等温情况一样,物理方程需计入温度变化的效应.例如

$$\varepsilon_{xx}=\frac{1}{E}[\sigma_{xx}-\nu(\sigma_{yy}+\sigma_{zz})]+\alpha T,$$

$$\varepsilon_{yz}=\frac{2(1+\nu)}{E}\sigma_{yz},$$

余类推.其中 α 是物体的线膨胀系数,T 是温度改变量,其他记号同一般弹性力学的记法.若以位移为基本变量求解问题,只需在弹性体的体力向量上附加向量

$$-\beta\left(\frac{\partial T}{\partial x},\frac{\partial T}{\partial y},\frac{\partial T}{\partial z}\right),$$

同时在应力边界附加面力 $\beta(l,m,n)$,式中 $\beta=(3\lambda+2\mu)\alpha$, l,m,n 是边界的外法向余弦,

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)},\quad \mu = \frac{E}{2(1+\nu)}$$

称为拉梅系数,其他关系式均同等温时一样.如果物体变形影响温度分布,则确定温度场时需在热传导方程中附加含应变的项,解题时再把热传导方程和热弹性关系联合起来.许多工程设备都涉及温度应力问题,故热弹性力学和固体力学的其他分支学科,如断裂力学、板壳力学、复合材料力学等,都有紧密的联系.由温度变化效应造成的结构失稳,称为热屈曲.

应力集中(stress concentration) 弹性力学术语.指构件因几何形状影响,使局部应力增高的现象.多见于孔洞边缘、凹缺角顶等处.这种应力与假定没有缺陷时该处应力的比值 k 称为应力集中系数,由几何形状以及载荷方式确定,与载荷大小无关.k 总大于 1,常常接近于 2,3,4,或更大.例如,受拉平板中间开一小圆孔,则孔边应力发生集中,$k=3$.若把圆孔改为长轴垂直拉伸方向的椭圆孔,则

$$k = 1 + 2\sqrt{\frac{a}{\rho}},$$

a 为长半轴,ρ 为该轴顶点的曲率半径,大的应力就在该处.可见椭圆愈扁,k 愈大,当 $a=25\rho$,则 $k=11$.若成为裂缝,k 将趋于无穷.对脆性材料,应力集中容易使其断裂;对塑性材料,将引起局部屈服,使应力重新分布而缓和,但在动荷情况则易疲劳而破坏.实用上,常对物件的缺裂凹角采取加强或消除尖角、增加光洁度等措施,以避其害.

应力集中系数(stress concentration coefficient) 见“应力集中”.

屈曲(buckling) 弹性力学术语.指弹性系统平衡解分岔的一种形式.令离散化后的弹性系统运动方程为

$$\dot{x} = F(x,\lambda),\quad x,F \in \mathrm{R}^n,\ \lambda \in \mathrm{R},$$

且令(x^*,λ^*)为平衡解,则平衡解的分岔条件相当于线性方程

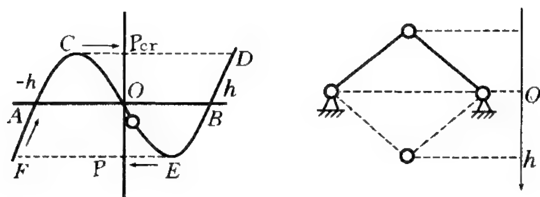
$$\left.\frac{\mathrm{D}F}{\mathrm{D}x}\right|_{(x^*,\lambda^*)}(x - x^*) = 0$$

中$(x-x^*)$有非零解,或由 $\|\mathrm{D}F/\mathrm{D}x\|=0$ 求解 λ^*,由上列非线性方程求得非零的解向量 $v=x-x^*$.这里 λ^* 称为临界参数,当 λ 表征荷载时称为临界荷载,v 称为屈曲形式.对于 λ 表征荷载时,一般最重要的是 λ^* 的最小值,称为第一临界载荷,较高的临界载荷和屈曲形式在实际工程设计中意义不大,但有时有理论上的兴趣.从发生屈曲的分岔点开始,产生新的平衡解分支,称为屈曲后平衡路径.在实际中需研究这个新分支上的平衡是否稳定.具体路径的几何状态称为屈曲后行为.研究结构屈曲后行为有现实工程意义,若某些结构在屈曲后,新的解分支是稳定平衡的,则结构物在承受超过临界荷载的外载时,并不丧失承载能力,根据这一特点可以充分利用

结构的承载力.

屈曲形式(buckling mode)　见"屈曲".

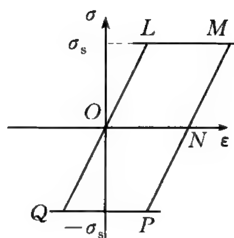
弹性突跳(snapping)　非线性弹性力学术语.指当参数微小变化时,弹性系统的平衡解从一个临界点到非邻近的一个稳定平衡点的动力跃迁.如图,两根铰接的弹性杆,在铅直力作用下,由于压力 P 的压缩,当 P 达到临界力 P_{cr}时,会动力跃迁到虚线的位置. A 点的高度,从初始 $-h$开始,当 P 达到 P_{cr} 时,w 会突然增大由 C 点达到 D 点,而当 P 反向加载时,又会从 E 点达到 F 点.平衡解曲线 $FACOEBD$ 的 COE 段平衡是不稳定的,在加卸载过程中总是沿着 $FACDBEF$ 的回路行进.这种现象是一个典型的弹性突跳.在复杂的弹性系统中也经常遇到弹性突跳现象,如弹性球壳在外压下的突然失稳,弹性拱在压力下的塌陷等.弹性突跳中临界点 C 和 E 处的结构失稳和屈曲不同,称为极值失稳.在一般极值失稳情况下,不一定存在另一非邻近的稳定平衡点,当外载达到临界载荷时会导致整个结构的破坏.这种失稳是结构设计中所不允许的.



极值失稳(extremum stable losing)　见"弹性突跳".

弹性系统稳定性的初缺陷理论(initial imperfect theory of stability of elastic system)　关于弹性结构稳定性的一种理论.是解释弹性结构稳定性问题中理论临界荷载与实验结果偏离较大,或实验结果比较分散的一种理论. 20 世纪 30 年代前后,人们已发现某些结构,如柱壳受轴压时,临界荷载只有理论结果的五分之一到二分之一. 20 世纪 30 年代初,人们开始提出关于柱壳初缺陷影响的猜测,到 1945 年,荷兰学者柯伊塔(Koiter,W. T.)在他的博士论文"关于弹性平衡的稳定性"中,开始引进了表征结构的初缺陷的参量,并讨论了临界荷载对初缺陷的依赖关系.在圆柱壳的情形,发现当外载为轴压时,有时临界荷载对于初缺陷变化是十分敏感的.由此开展了这方面的研究,并成功地解释了实验临界荷载分散的问题.

理想塑性(perfectly plastic behaviour)　变形体的一种力学模型.是忽略材料加工硬化性质的固体力学模型,可分为弹塑性和刚塑性两类.前者弹性变形和塑性变形量级接近;后者弹性变形比塑性变形小得多,可以不计.理想弹塑性假定应力应变的线性阶段之后是完全塑性流动,拉伸和压缩完全对称,塑性变形不影响弹性性质.可用简单拉伸压缩的应力 σ 和应变 ε 的坐标图说明.如图,σ_s 是屈服应力,直线 LM 和 PQ 都平行于 ε 轴,LOQ 的斜率等于杨氏模量 E,$LO=OQ$.在加载卸载和再加载等任意过程中,σ-ε 图象只能是平行于 LOQ 的直线(含 LOQ),大小为上至 LM 下至 PQ(含延长线).任意可能的(ε,σ)总在这类折线上.例如从自然状态拉伸时,则沿 OL 变动,当 $\sigma\geqslant\sigma_s$,则沿 LM 流动.若从 M 卸荷,则沿 MN 到 N 留有塑性变形 ON.若再拉伸,则沿 NM 变化.反之,若压缩,则沿 NP 进行,当压应力变为 $-\sigma_s$,则沿 PQ 流动……若令线段 LM 和 QP 都水平移动使点 L 和 Q 到 σ 轴就成为刚塑性的 σ-ε 图,相当于 $E\to\infty$的情况,即材料在 $\sigma<\sigma_s$ 时不发生变形,$\sigma\geqslant\sigma_s$ 却塑性流动.这些模型都突出了问题的核心,简化了计算,便于建立塑性力学的理论而指导实践,例如在结构分析、金属加工等方面得到广泛的应用.



理想刚塑性(rigid-perfectly plastic behaviour)　一种变形体模型.在弹塑性变形过程中,略去弹性变形且不考虑塑性应变增大时,应力的改变得到的应力应变关系.它是金属等材料本构关系的一种近似.按照这种本构关系,当应力未处于屈服面上时,应变为零,当应力处于屈服面上时,产生塑性流动.由于这个简化的本构关系使计算大大简化,所以它在计算结构的极限承载能力(亦称结构极限分析)和判定安定性以及金属塑性成形的力学分析中得到广泛的应用.

沙堆比拟(sand heap analogy)　杆件塑性扭转时求扭矩 M 的比拟法.干沙成堆,表面斜度 $\tan\alpha$ 必等于内摩擦系数,为常数.杆件塑性扭转横截面最大切应力 τ_m 必达屈服值 K.根据扭转理论,取坐标系 $Oxyz$,令 x 轴、y 轴在杆的横截面内,设应力函数 $\varphi(x,y)$在边界为 0,在面内

$$\frac{\partial\varphi}{\partial x}=-\tau_{zy},\quad \frac{\partial\varphi}{\partial y}=\tau_{zx},$$

则满足平衡方程,由其他条件有

$$2\iint\varphi\mathrm{d}x\mathrm{d}y=M$$

及

$$\left|\frac{\mathrm{d}\varphi}{\mathrm{d}n}\right|=\tau_m\ (\text{等于}\ K),$$

n 是曲线 $\varphi=c$ 的法向.这表明 $z=\varphi(x,y)$图象之斜度为常数 K,与沙堆有共性.只要使沙堆底面与杆的横截面相同,则存在相似关系.沙体积用 V 表示,

则有

$$V : \iint \varphi \mathrm{d}x\mathrm{d}y = \tan\alpha : K,$$

故 $M=2KV/\tan\alpha$. 此法适用于杆截面为单连通情形.

残余应力(residual stress)　一类应力. 指固体在外部作用(荷载、温度等)下，由于内部发生了塑性变形，当外部作用除去后，内部所遗留的应力. 残余应力有时会使结构物承载能力减小，甚至引起破坏，有时又会使结构物增加承载能力. 计算残余应力的方法原则上要考虑塑性变形，因而是塑性力学的课题. 实践上多采用实测办法，一般将有残余应力的局部打孔，然后测量孔的变形，利用弹性力学理论反算局部应力.

安定性理论(shake-down theory)　涉及塑性变形过程的一种理论. 是塑性力学的一类关于在交变外载下塑性变形是否积累的理论. 在外载的可变部分作用下，如果物体的塑性变形逐渐积累，或经变化往复后，使局部应变的幅度不停止地增大，则认为这个系统是不安定的，在多次交变荷载时，可能导致系统破坏. 否则，在多次交变荷载作用下，系统达到某一应变后不再增大，就说系统是安定的. 使用安定性理论校核结构强度，可以使结构更加可靠.

材 料 力 学

梁(beam)　一种构件. 是细长可变形柱体，且外力垂直于母线. 在这种外力作用下，梁变形主要体现为弯曲，它可用梁的轴线的位移来描述，称为挠度. 梁的轴线在弯曲时不产生伸缩变形，称为中性轴. 在线性弹性的条件下，梁截面又具有对称性，而且外力作用在截面的对称线上时，梁的挠度 $w(z)$ 满足

$$EJ\frac{\mathrm{d}^2 w}{\mathrm{d}z^2} = M(z),$$

这里 E 为杨氏模量，z 为中性轴，$M(z)$ 为梁的弯矩，J 为截面惯性矩：

$$J = \iint_F y^2 \mathrm{d}F,$$

F 为梁截面. 弯矩与剪力 $Q(z)$、分布力 $q(z)$ 的关系为

$$-\frac{\mathrm{d}Q}{\mathrm{d}z} = q(z),\ \frac{\mathrm{d}^2 M}{\mathrm{d}z^2} = q(z),\ \frac{\mathrm{d}M}{\mathrm{d}z} = -Q(z).$$

令梁长为 l，则在 $z=0$，l 两端各提两个边条件，即在挠度 w、转角 $\mathrm{d}w/\mathrm{d}z$、弯矩 M、剪力 Q 四个量中每端给定两个，在总共四个边条件中位移和转角不得少于两个. 在这些边条件下，梁的挠度可以积分出来. 工程中许多构件都可以近似地看成梁. 梁的线性理论是工程力学和材料力学中的重要课题.

铁木森柯梁(Timoshenko beam)　亦称深梁. 一种梁的理论. 是计入剪切对弯曲影响的梁. 铁木森柯(Timoshenko, S. P.)于 1921 年首先建立其理论. 梁是受弯曲的杆件，一般的梁不计入剪切和挤压的影响，而服从平截面假定：变形前垂直于梁轴的截面，变形后仍垂直，适应于高跨比不大于 1/8～1/5 的情形. 随着高跨比的增大，剪切和挤压的影响也相应增大，而剪切尤甚. 铁木森柯的理论就考虑了剪切的影响，把平截面假定下的轴向位移公式

$$u = -z\frac{\mathrm{d}w}{\mathrm{d}x},$$

修正为

$$u = -z\frac{\mathrm{d}w}{\mathrm{d}x} + \gamma z,$$

式中 $\gamma=\gamma(x)$. 如图所示，x 为沿梁的轴，u，w 分别为梁在 x，z 方向的位移，γ 为梁的横向剪切角，即 x 轴与 z 轴角度的改变量. 这就能适应于高跨比不大于 1/2.5～1/2 的情况. w 和 γ 的微分方程为

$$EIw^{\mathrm{IV}} = q(x) - \frac{EI}{GA}q''(x),$$

和

$$\gamma''' = -\frac{q''}{GA},$$

法应力

$$\sigma_{xx} = E[-w''z + \gamma' z].$$

式中 E 为杨氏模量，I 为梁的截面对中性轴的惯性矩，$q(x)$ 为载荷密度，G 为剪切模量，A 为截面面积. 这是把剪切效应沿截面平均化了. 为消除这种误差，又要保持一维计算的简便，人们又引入因子 K(有的取为截面的函数，有的取为泊松比 ν 的函数)，令切应力 $\tau=KG\gamma$，这时上面方程中含 G 的地方均需换为 KG. 当高跨比进一步增大，则挤压影响也增大，弯曲的中性层将明显地离开几何中位，稳定性问题也必突出，整个问题需要重新考虑.

连续梁(continuous beam)　多支座梁. 该梁具有三个或三个以上支座，以承受侧向荷载的连续直杆结构. 增加中间支座减少梁的跨度，能显著提高梁的刚度和强度，如局部荷载作用下的连续梁，其内力较静定梁分布更趋均匀，挠度有明显减小. 连续梁多用于桥梁、房梁、架空的长管道和精密机床主轴等，它为超静结构，计算较静定梁复杂，单用平衡条件不能求解，必须同时考虑变形协调条件才能求解，如力法、位移法、力矩分配法、矩阵位移法和三弯矩方程等，当今工程中多用力矩分配法和矩阵位移法.

弹性地基梁(beam on elastic foundation)　一种受特别约束的梁.是支承在弹性地基上受有侧向荷载的杆件结构.多见于房屋建筑中的长条基础.为解决该结构与地基间的力的传递,通常采用温克勒假设:梁的每点挠度 w 等于该点地基的沉陷量,地基给该点的反力与该点的沉陷量成正比,表示为 Kw,K 为基床系数,单位面积地基被压陷单位深度所需加的力,与附近各点的反力无关.由此得梁的挠度方程为

$$EI_y\frac{d^4w}{dx^4}+Kw=q,$$

式中 E 为梁材料的弹性模量,I_y 为梁截面对水平轴的形心主惯性矩,q 为作用在梁上的分布外荷载.对于无限长梁,中间受集中力 p 作用,梁的挠度表达式为

$$w=\frac{p\beta}{2k}e^{-\beta x}(\cos\beta x+\sin\beta x),$$

式中 $\beta=(K/4EI_y)^{1/4}$,x 为某点至 p 作用点的距离(左、右相同).梁中弯矩为

$$M=-EI_y\frac{d^2w}{dx^2}$$

$$=-\frac{p}{4\beta}e^{-\beta x}(\sin\beta x-\cos\beta x).$$

最大挠度 w_{max} 和最大弯矩 M_{max} 出现在外载 p 作用点,分别为

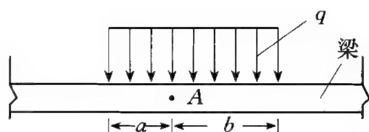
$$w_{max}=\frac{p\beta}{2K},\quad M_{max}=\frac{p}{4\beta}.$$

当在无限长梁中的一部分有分布外荷载 q 作用时,利用上述结果并通过积分可求得某点 A(见图)处的挠度和梁中弯矩:

$$w=\frac{q}{2K}(2-e^{-\beta b}\cos\beta b-e^{-\beta a}\cos\beta a),$$

$$M=\frac{q}{2K}(e^{-\beta a}\sin\beta a+e^{-\beta b}\sin\beta b),$$

式中 a,b 如图所示.考虑到端部条件,即可得到有限弹性地基梁的挠度和弯矩.对一般的有限长弹性地基梁通常采用数值方法求解微分方程,如有限单元法、有限差分法.



梁的挠度(deflection of beam)　梁变形的主要表征.梁的横向位移,变形后的梁轴线为挠曲线.引起变形的因素有侧向荷载、温度的变化、支座移动等.对于工程中常用的梁,其挠曲线方程与分布荷载 $q(x)$ 之间的微分关系为

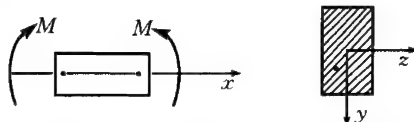
$$\frac{d^4w}{dx^4}=-\frac{q(x)}{EI},$$

式中 $w=w(x)$ 为挠度,E 为材料的弹性模量,I 为截面惯性矩.EI 反映梁抵抗弯曲变形的能力,称为抗弯刚度.该关系式是根据小挠度理论推出的,即当

$$\frac{dw}{dx}\ll 1$$

时,有足够的精度.在给定边值条件下,积分上式可得出挠曲线方程表达式,同样也可用数值方法、变分原理方法求解.

梁的纯弯曲应力(pure stress of bending in beam)　梁中的一种应力.指梁上只有外力偶矩 M 作用的梁截面上的正应力.当 M 所在的平面垂直于



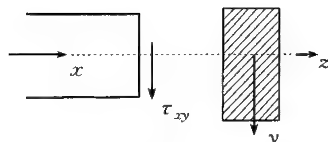
截面一根主惯性轴时,梁截面上仅有弯矩而无剪力的弯曲,称为纯弯曲.该梁变形后,原截面仍为平面(即为平面假设),且垂直于变形后的轴线,称为平面弯曲.梁横截面转动,使梁凹边的纤维缩短,凸边纤维伸长.中间有一层纤维,无长度改变的层,称中性层.该层与横截面的交线为中性轴.梁横截面上的正应力 $\sigma=My/I_z$,y 为求应力所在的点到中性轴之距离,I_z 为截面对 z 轴(见图)的主轴惯性矩.应力沿截面为线性分布.

梁的纯弯曲(pure bending of beam)　见"梁的纯弯曲应力".

梁的应力(stress in beam)　梁在弯曲时的应力.指在横向荷载作用下,梁截面上的应力(正应力和剪应力),称为梁的横向弯曲应力.梁的弯曲分为纯弯曲、横向弯曲和斜弯曲三种.当有横向荷载时,梁的横截面上不仅有正应力 σ,而且还有剪应力 τ.由于剪应力的存在,梁的横截面将发生翘曲.此外,在与中性层平行的纵截面上有挤压应力.因此,纯弯曲中的平面假设和纵向纤维间互不挤压假设都不能成立.但在工程中,当梁(如矩形截面)的高跨比小于1/5时,横截面上的正应力仍可按纯弯曲应力公式计算.剪应力 τ_{xy} 为

$$\frac{Q_yS_z}{I_zb},$$

式中坐标如图所示;Q_y 为 y 方向剪力;S_z 为所求剪应力所在的点以外部分的截面对 z 轴的面积矩;b 为梁在 z 轴方向的宽度.通常最大正应力发生在距 y 轴最远处,而最大剪应力发生在 z 轴处.当横向力产生的弯矩所在平面并非垂直于截面中任一根主惯



性轴时的弯曲为斜弯曲.将弯矩分解为分别垂直于

两个主惯性轴的弯矩 M_y 和 M_z（见图），斜弯曲引起的正应力可由两个弯矩分量分别求出，然后叠加，

$$\sigma_x = \frac{M_z y}{I_z} + \frac{M_y z}{I_y},$$

式中，y 和 z 分别为所有应力所在的点到 z 轴和到 y 轴的距离；I_y 和 I_z 分别为截面对 y 轴和 z 轴的主惯性矩. 通常细长实心梁横截面上的最大剪应力远小于最大正应力，可以忽略不计. 但对于薄壁梁和夹心梁，则必须考虑剪应力影响.

梁的横向弯曲(crosswies bending of beam) 见“梁的应力”.

梁的斜弯曲(oblique bend of beam) 见“梁的应力”.

梁的剪切效应(shearing effect of beam) 剪力对梁变形的影响. 指计入剪应力对梁挠度的影响. 根据变形体虚功原理，可得出在荷载作用下计算梁的某点挠度 Δ_{ip} 的公式：

$$\Delta_{ip} = \int_L \frac{\overline{M}M_p}{EI}dx + \int_L \frac{k\overline{Q}Q_p}{GA}dx,$$

式中 M_p 和 Q_p 分别为荷载引起的梁内弯矩和剪力；$\overline{M}$ 和 $\overline{Q}$ 分别为单位荷载法引起的梁内弯矩和剪力；E 和 G 分别为材料的弹性模量和剪切模量；I 和 A 分别为梁截面惯性矩和面积；L 为区间长度；式中前一项积分为弯曲引起的挠度，后一项积分为剪力引起的挠度；k 为截面形状系数，如矩形截面为 6/5，圆形截面为 10/9，薄壁圆管为 2，工字形或箱形截面为 A/A_1，A 为截面总面积，A_1 为腹板面积. 简支梁跨度为 L，受均匀分布载荷 q，其梁中点挠度为

$$\Delta_{中} = \frac{5ql^4}{384EI}\left(1 + \frac{48kEI}{GAl^2}\right),$$

式中等号右方第一项为弯曲产生的挠度；第二项为剪力对挠度的影响. 若梁为矩形截面，$k=1.2$，$I/A=h^2/12$（h 为截面高度），设泊松比 $\nu=0.3$，则

$$\frac{E}{G}=\frac{8}{3},$$

当梁的高跨比为 1/10 时，剪切变形影响约为弯曲变形影响的 2.56%. 对于一般梁，可略去剪切变形对位移的影响. 但当梁高 $h>l/5$ 时，剪切变形与弯曲变形对位移的影响大于 10.2%，即对深梁或承剪能力差的梁，则必须考虑剪切对挠度的影响.

弯曲(bending) 材料力学术语. 是细长杆件或薄板结构在侧向荷载或温度变化等因素作用下，其杆轴线或板中面曲率发生变化现象（参见“挠度”）.

荷载(load) 亦称载荷. 力学概念. 一般指主动作用于结构的外力及物体间相互作用力. 温度变化、基础沉陷、材料收缩等也能引起物体的变形和内力，故也称荷载. 按作用位置分为体积力或惯性力、面力（如水压力）、线力（如绳索缠绕力）、集中力. 根据荷载作用的久暂可分为恒载和活载. 恒载是长期作用在结构上的不变荷载，如结构自重、土压力等；活载是其值随时间变化的荷载，如楼面和屋面活载、吊车荷载、雪载和风载等. 根据作用的性质分为静力载和动力载. 静力载的数量、方向、位置均不随时间变化或变化极为缓慢，不使结构产生明显的加速度；动力载是随时间迅速变化或在短时间内突然作用或消失的荷载，使结构产生显著的加速度$\frac{\partial^2 y}{\partial t^2}$而引起的惯性力

$$F = m\frac{\partial^2 y}{\partial t^2},$$

不能忽略，其中 y 为质点的位移，m 为质量. 动力载又分周期载（如简谐荷载 $p(t)=p_0\sin\theta t$，式中 p_0 为荷载幅值，θ 为干扰力频率）、一般周期载及冲击载，如爆炸型荷载

$$p(t) = \begin{cases} p_0\left(1 - \frac{t}{t_0}\right) & (t \leqslant t_0), \\ 0 & (t > t_0), \end{cases}$$

式中 p_0 为初始值，t_0 为荷载持续时间，t 为时间坐标和随机荷载，如地震力等.

载荷(load) 即“荷载”.

荷载效应组合(combination of load effect) 结构设计时假定的荷载状况. 考虑极限状态确定其荷载效应时，所有可能同时出现的诸荷载的组合. 组合后的效应为总效应. 荷载效应组合的设计值 S 乘以结构重要性系数 γ_0 后，应小于或等于结构的抗力值 R，即 $\gamma_0 \cdot S \leqslant R$. 组合分为基本组合和偶然组合. 不考虑偶然荷载的组合称为基本组合. 根据不同设计要求，又可分成短期效应组合 S_s 和长期效应组合 S_l. 基本组合的设计值 S 按下式确定：

$$S = \gamma_G C_G G_k + \gamma_{Q1} C_{Q1} Q_{1k} + \sum_{i=2}^{n} \gamma_{Qi} C_{Qi} \psi_{ci} Q_{ik},$$

式中 γ_G 为恒载分项系数，取 1～1.2；γ_{Q1} 和 γ_{Qi} 分别为第 1 个和第 i 个活荷载的分项系数，一般取 1.4；G_k，Q_{1k} 和 Q_{ik} 分别为恒载的标准值、第 1 个和第 i 个活荷载的标准值；C_G，C_{Q1} 和 C_{Qi} 分别为恒载、第 1 个和第 i 个活载的荷载效应系数；ψ_{ci} 为第 i 个活荷载的组合值系数，有风载时取 0.6，无风载时取 1.0. 短期和长期效应组合为：

$$S_s = C_G G_k + C_{Q1} Q_{1k} + \sum_{i=2}^{n} C_{Qi} \psi_{ci} Q_{ik},$$

$$S_l = C_G G_k + \sum_{i=2}^{n} C_{Qi} \psi_{qi} Q_{ik},$$

式中 ψ_{qi} 为第 i 个活荷载标准永久值系数. 上面提出的荷载效应系数，均可由各种活荷载标准值查得.

拱(arch) 一种构件. 是杆轴线为曲线，在竖向荷载作用下产生水平推力的杆件. 中国拱在桥梁上

的应用始于晋朝，另外还多用于房屋、遂洞等. 由于水平推力(或拉杆内力)使梁内弯矩大大减小，为此，可使用抗压性能较好的材料(如石料、砖、混凝土等)制成拱结构. 工程中常用拱的形式有：三铰拱(图1)、两铰拱(图2)、无铰拱(图3)和有拉杆的三铰拱(图4)与两铰拱(图5). 三铰拱为静定结构，水平推力 H 可用平衡条件求出，

$$H=\frac{M_c^0}{f},$$

式中 f 为拱高，M_c^0 为同等跨度代梁 c 截面弯矩值；

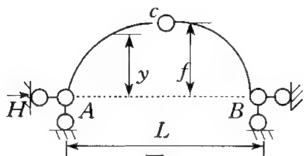


图1

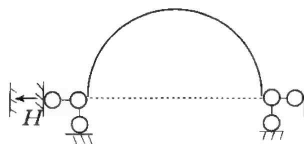
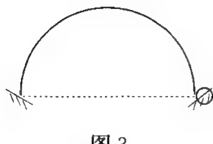
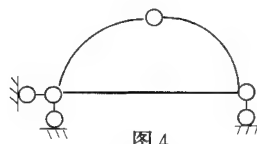
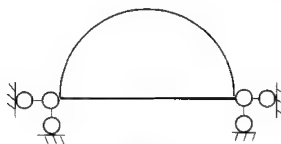


图2


图3


图4


图5

两铰拱为一次超静定结构，其水平推力 H 则由变形条件求出，拱内弯矩

$$M=M^0-Hy,$$

式中 M^0 为同等跨度代梁在同样荷载作用下所求截面弯矩值，y 为截面形心到两底铰连线垂直距离(图1)；无铰拱为三次超静定结构，可用力法求解. 拱内截面上弯矩、剪力和轴力一般可以合成作用在截面上的总压力，各截面上的总压力作用点的连线称为拱的压力线，当压力线与拱轴线重合时，则该轴线为合理拱轴，该拱截面上仅有均匀分布的轴向压力. 无铰拱弯矩分布较均匀，但支座的较少量移动对拱内应力分布有较大影响，故较差的地基不易采用. 作为屋盖的拱，常采用有拉杆的拱结构.

梁柱(beam-column)　亦称压弯杆. 一种构件. 是同时承受轴心压力和横向载荷或承受偏心压力的细长杆. 在 xz (x 为轴向，z 为截面的形心主惯性轴)平面内受有横向载荷 q 和轴向压力 N_x 共同作用，沿 z 轴方向挠曲 w 的微分方程为

$$EI_y\frac{\mathrm{d}^4w}{\mathrm{d}x^4}+N_x\frac{\mathrm{d}^2w}{\mathrm{d}x^2}=q,$$

式中 E 为材料的弹性模量，I_y 为截面对形心主惯性 y 轴的惯性矩. 当梁柱受偏心压力 N_x，在 xz 平面引起弯矩 M_y 或直接承受轴心压力及外力矩 M_y 作用下，它的挠曲微分方程为

$$EI_y\frac{\mathrm{d}^2w}{\mathrm{d}x^2}+N_xw=M_y,$$

当 N_x 远小于等截面细长压杆的临界力 p_{cr} 时，则计算侧向挠度不需要考虑轴向力的影响. 计算应力，只需分别求出轴力和弯矩引起的应力后，叠加即可. 若 $N_x\leqslant0.6p_{cr}$ 时，则梁柱挠度 w、弯矩 M 和正应力 σ 的近似计算公式为

$$w=cw_0,\quad M=cM_0,\quad \sigma=\frac{N_x}{A}+c\sigma_0,$$

其中 w_0，M_0 和 σ_0 分别为无轴力时梁柱的挠度、弯矩和正应力；

$$c=\frac{1}{1-N_x/p_{cr}}$$

为放大系数；A 为截面面积. 当 $N_x=0.6p_{cr}$ 时，上面的近似计算误差小于2%. 若 N_x 接近于 p_{cr}，则需直接求解挠曲微分方程，再求应力.

压弯杆(compressive-bending bar)　即“梁柱”.

临界压力(critical pressure)　亦称临界力. 材料力学概念. 指中心受压等截面直杆所能承受压力的限界值. 临界压力 p_{cr} 的表达式最早由瑞士数学家欧拉(Euler, L.)导出：

$$p_{cr}=\frac{\pi^2EI}{(\mu l)^2},$$

其中 E 为材料的弹性模量；I 为截面的主惯性矩；EI 表示杆件抵抗弯曲的能力；l 为杆长；μ 为不同杆端约束下临界力的折减系数，如两端铰支座时 $\mu=1$，一端固定、一端铰支座时 $\mu=0.7$，一端固定、一端自由时 $\mu=2$，两端固定时 $\mu=1/2$. 欧拉压杆临界应力

$$\sigma_{cr}=\frac{p_{cr}}{A}=\frac{\pi^2EI}{(\mu l)^2A},$$

A 为截面面积，或写成

$$\sigma_{cr}=\frac{\pi^2E}{\lambda^2},$$

其中 $\lambda=\mu l/i$，$i=(I/A)^{\frac{1}{2}}$. λ 称为压杆细长比或压杆的柔度. 当 λ 较大而 $\sigma_{cr}<\sigma_p$(σ_p 为比例极限)时，称为长压杆. 若 λ 减小使 $\sigma_p<\sigma_{cr}<\sigma_b$($\sigma_b$ 为强度极限)，则称为中压杆(或称长柱)，此时应以切线模量 E_t(对应于 σ_{cr} 处的 $\mathrm{d}\sigma/\mathrm{d}\varepsilon$，$\sigma$ 和 ε 为应力和应变)代替式中的 E. 又由于 σ_{cr} 与 E_t 相互关联，要用逐次近似法求 E_t 和 σ_{cr}. 若 λ 再减小使 $\sigma_{cr}>\sigma_b$，则称短压杆，短压杆

破坏前不出现屈曲.工程上计算压杆极限承载应力的简化式为

$$\sigma = \frac{1+\gamma}{1+\gamma+\gamma^2}\sigma_b,$$

其中 $\gamma=\sigma_b/\sigma_{cr}$.

临界力(critical pressure)　即"临界压力".

压杆稳定(stabilization of compressive bar)　杆件受压力时的平衡失稳现象.中心受压直杆在临界力 p_{cr}作用下,其直轴线形状下的平衡丧失稳定性,又称失稳.理想细长直杆,当压力增至临界力 p_{cr}时,在很小的侧向扰动后,却可能出现很大的侧向位移,甚至失去承载能力,即是压杆的失稳现象,称压杆的屈曲.

压杆的柔度(flexibility of compressive bar)　见"临界压力".

挠度(deflection)　材料力学概念.指细长杆件或板、壳等变形后的侧向位移.其原因是受侧荷载或温度变化等因素的影响.细长杆挠度是指杆轴线上各点在垂直杆轴线的位移量;板壳挠度即为中面(等分厚度各点而组成的面)上各点垂直于原中面的位移量.当物体几何尺寸及约束条件一定后,物体各点随位置及荷载而变化的位移可用位移函数或挠度函数来表示.如对于静力问题,挠度函数表示为

$$w=w(x,q) \text{ 或 } w=w(x,y,q),$$

式中 $q=q(x)$或 $q=q(x,y)$.对于薄板问题,挠曲面偏微分方程为

$$\frac{\partial^4 w}{\partial x^4}+2\frac{\partial^4 w}{\partial x^2\,\partial y^2}+\frac{\partial^4 w}{\partial y^4}=\frac{q(x,y)}{D},$$

式中 D 为板的抗弯刚度,w 为板的静挠度.对于动力问题,侧向荷载除是位置的函数外,还是时间 t 的函数,$q=q(x,t)$或 $q=q(x,y,t)$,其挠度 $w=w(x,t)$或 $w=w(x,g,t)$.在给定边界条件情况下,可用积分法、变分原理以及近似解法求出挠曲面表达式,或指定点的挠度值.

疲劳(fatigue)　材料力学术语.指在交变应力下材料局部产生损伤积累直至整体破坏的现象.由它造成的断裂称疲劳破坏.材料的疲劳是由法国数学家、力学家彭赛列(Poncelet,J.-V.)于 1839 年首先研究,19 世纪 50～60 年代,法国科学家沃勒(Wöhler,A.)做进一步研究,从而奠定了它的基础.疲劳研究主要是对不同材料通过实验给出 S-N 曲线,这里 S 表示应力水平,N 表示在这一应力水平下达到破坏的循环次数,亦称疲劳寿命.近代研究中,将疲劳现象同微裂纹的扩展和产生联系起来,并研究疲劳同环境条件、材料工艺过程以及成分的关系.疲劳研究的结果可为交变载荷下结构设计提供依据.

损伤(damage)　材料力学术语.指材料或构件内部微小缺陷的累积过程.微小缺陷指分布在材料内部的错位(晶体中分子排列的错位)、微裂纹、微空洞等.研究损伤问题,目前多采用新的标量、向量或张量来描述缺陷连续化,这些标量、向量或张量称为损伤变量,然后利用统计的和连续介质力学的办法得到这些变量同应力张量、应变张量之间的关系,称为损伤本构关系,进而求解损伤变量和应力、应变.利用求得的损伤变量可以评价材料强度、寿命和韧性.对损伤规律的研究,目前是材料科学和固体力学比较活跃的方向,有时被称为损伤力学.

许用应力(allowable stress)　亦称资用应力.对应力的限制.保证受力结构安全(不破坏、不失稳、不因变形损害应有的功能)所许可的最大应力,记为

$$[\sigma]=\frac{\sigma^0}{k}.$$

σ^0 为该材料破坏的危险应力,对静荷下的塑性体为屈服应力,脆性体为断裂时的应力(称强度极限),在交变应力下为疲劳极限.k 为大于 1 的安全系数,因材料本身和人工处理条件以及使用环境和目的的不同而异,工程上有具体规范.可依强度理论和力学公式求得任意应力状态的相当应力 σ_{eq},实用上令 $\sigma_{eq}\leqslant[\sigma]$,即能保证结构安全.

资用应力(allowable stress)　即"许用应力".

强度理论(strength theory)　任意应力状态下材料破坏的判据理论.破坏,对塑性材料指屈服,对脆性材料指断裂,塑性材料屈服后再断裂称为切断.破坏和物体形状、性质、工作条件、受力状态等情况有关,相当复杂.现有理论限于考虑常温和静荷,并认为危险点的破坏只取决于该点的应力状态,和邻近的应力无关.设 $\sigma_1\geqslant\sigma_2\geqslant\sigma_3$ 为某点的三个主应力,σ^0 为单向拉伸材料破坏的危险应力.各理论为:

1. 最大拉应力理论.最大拉伸应力达危险值,材料即破坏.条件为 $\sigma_1\geqslant\sigma^0$,适应于拉应力为主的脆性材料.1638 年,意大利科学家伽利略(Galilei,G.)认为材料破坏是因拉伸或压缩最大应力达危险值.此观点并非全对,修正如上.

2. 最大伸长变形理论.最大伸长应变达到危险值,材料即破坏.经折算,条件可表示为相当应力,即

$$\sigma_{eq}=\sigma_1-\nu(\sigma_2+\sigma_3)\geqslant\sigma^0,$$

ν 为泊松比,适应于脆性材料.1682 年,马里奥特(Mariotte,E.)提出时,认为最大伸长或最大缩短的应变都是破坏原因.此观点也非全对,修正如上.

3. 最大切应力理论.最大切应力达到危险值,材料即破坏.条件为相当应力,即

$$\sigma_{eq}=\sigma_1-\sigma_3\geqslant\sigma^0.$$

不考虑中间主应力 σ_2 的影响,适应于拉伸和压缩屈服应力大小相等的塑性材料.实用上,按此理论设计的尺寸较为富余.困难在于不能解释三向拉伸遭破

坏的实际，对某些单向拉压危险应力值不同的材料也不能适应. 1773 年，库伦(Coulomb,C. A. de)提出时所指的破坏是被切断，后来特雷斯加(Tresca,H.)明确为塑性流动，成为塑性力学中的一个屈服条件.

4. 歪形能理论. 材料破坏是因总的变形比能中形状改变的那部分比能，即歪形比能达到危险值. 该条件可折算为相当应力，即

$$\sigma_{eq}=\frac{1}{\sqrt{2}}\sqrt{(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2}\geqslant\sigma^0.$$

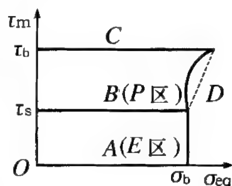
某点处的歪形比能正比于该点的统计平均切应力，故此理论也是一种切应力理论. 适合于塑性屈服条件，较上述最大切应力理论精确，能发挥材料富余的强度储备. 1885 年，意大利学者贝尔特拉米(Beltrami,E.)曾设想总变形比能促使材料破坏，但和实验结果不符. 1904 年，波兰学者胡伯(Huber,M. T.)修改为此理论. 1913 年，德国数学家、力学家米泽斯(Mises,R. von)也如此提出，成为塑性力学中又一个屈服条件，故也称为胡伯-米泽斯屈服条件.

5. 莫尔强度理论. 1900 年，莫尔(Mohr,O.)在盖斯特(Gest,J. J.)的试验基础上，考虑材料受切力滑移时内摩擦力和正压力的关系，认为滑移面上法应力必影响屈服，发展了最大切应力理论，经适当简化，得到屈服条件为

$$\sigma_{eq}=\sigma_1-\frac{\sigma^0}{\sigma^0_{压}}\sigma_3\geqslant\sigma^0,$$

式中 $\sigma^0_{压}$ 为该材料的压缩破坏应力，取绝对值. 这克服了上述最大切应力理论在 $\sigma^0\neq\sigma^0_{压}$ 时的困难，但同样未考虑中间的主应力 σ_2 的影响，也不适应于断裂破坏.

6. 联合强度理论. 1946 年，苏联学者达维靖可夫(Давиденков,Н. Н.)和弗里特曼(Фридман,Я. Б.)基于同一材料可因应力状态不同而呈现脆性或塑性的事实，认为破坏与材料性质和受力条件都有关，所以，将“最大伸长变形理论”与最大切应力理论”结合起来，建立了该理论. 取 $\sigma_{eq}=\sigma_1-\gamma(\sigma_2+\sigma_3)$ 和 $\tau_m=\sigma_1-\sigma_3/2$ 为坐标轴，令 τ_s，τ_b 分别代表屈服和切断的最大切应力，σ_b 为单向拉断的应力，作屈服、切断和断裂三直线分别如图中 B，C，D，其位置因材料不同而异. 该理论可由图说明，当应力 σ_3，σ_2，σ_1 按比例由小到大变化，就形成过原点的直线 A. 若直线先通过断裂线 D，就表示这种应力组合使材料以断裂而破坏，若直线先通过屈服线 B 再通过切断线 C，就表示该应力组合使材料先屈服再切断. 平面上点的位置反映材料的应力状态，图中 E 区、P 区、切断线之上和断裂线之右分别代表该应力状态使材料安全、屈服、切断和断裂. 当 σ_{eq} 为负值，就以 τ_m 轴代表，说明这时材料只能剪切破坏. 可见脆性材料的断裂线必靠近 τ_m 轴，塑性材料的断裂线则远离 τ_m 轴. 该理论表明了应力状态对材料呈塑性或呈脆性具有影响，给金属加工技术提供了理论指导，不过也具有最大伸长变形理论和最大切应力理论具有的缺点.



结构力学

结构静力学(structural statics)　结构力学的一个分支. 它研究杆件结构在静荷载、温度变化和支座移动等因素作用下的内力、位移、承载能力和优化等问题. 其任务是对工程结构设计提供可靠的静力学理论分析方法和计算公式，求出结构中的内力、应力、应变和变形，以判断结构设计的合理性并提出优化设计，研究新型结构. 其内容包括理论分析和实验研究两个方面，对工程中所用的一切结构均应作静力学分析.

静定结构(statically determinate structure)　一种类型的结构体系. 指无多余约束的几何不变体系. 其内力及支反力可由平衡条件完全确定. 该结构的基本静力学特性是满足平衡条件内力解的惟一性. 在此基础上派生出如下特性：

1. 温度变化、支座移动和制造误差等因素不引起该结构内力.

2. 结构的局部平衡性.

3. 结构的荷载等效性.

4. 结构的构造变换性.

超静定结构(hyperstatic structure)　亦称静不定结构. 一种类型的结构体系. 指其内力、支反力不能仅从静力平衡条件求出，而必须同时考虑变形协调条件才能求解的结构. 其解法有两大类：

1. 力法类型(包括力法、余能法、矩阵力法、渐近力法).

2. 位移法类型(包括位移法、势能法、矩阵位移法、力矩分配法、无剪力分配法).

超静定结构的特性为：

1. 存在多余约束(仅由静力平衡方程不能求解的未知量)，结构刚度和稳定性有所提高.

2. 各杆刚度改变对内力分布有影响，即在荷载作用下，其内力分布与各杆刚度比值有关，而与其绝对值无关.

3. 温度变化、支座移动、制造误差均引起内力和变形.

静不定结构(statically indeterminate structure)　见“超静定结构”.

几何不变体系(geometrically invariable sys-

tem) 一种具有某类几何不变性的结构体系. 是在不考虑材料应变的条件下，体系的几何形状和位置都是不可改变的体系. 一般结构都必须是几何不变体系，不能采用几何可变体系. 无多余约束的几何不变体系为静定结构，仅用静力平衡方程：$\sum x=0$，$\sum y=0$和$\sum M=0$可求出全部内力及支反力. 式中各$\sum=0$分别表示作用于体系上的所有力对坐标轴的投影和对某轴的矩均为零. 有多余约束的几何不变体系为超静定结构.

几何可变体系(geometrically variable system) 不具几何不变性的结构体系. 指在不考虑材料应变的条件下，体系的位置或几何形状是可以改变的体系. 该体系不能作为结构.

刚架(rigid frame) 一种类型的结构. 指由全部刚结点或部分刚结点组成的直杆结构. 若所有杆件或荷载均在同一平面内，则为平面刚架；否则为空间刚架. 承载后，刚架发生变形，但在刚结点处，各杆不能发生相对移动和转动，即各杆间的夹角始终保持不变. 刚结点可以承受和传递轴力、剪力和弯矩. 平面刚架中各杆一般产生轴力、剪力和弯矩三种内力；空间刚架中各杆则产生一个轴力、一个扭矩、两个互相垂直的弯矩和剪力等六个内力. 工程中常用的刚架，大部分为空间超静定刚架，特定条件下可简化为平面刚架. 常用力法、位移法、迭代法、矩阵力法和矩阵位移法等方法求解.

桁架(truss) 一种类型的结构. 指全部由铰结点连结的直杆、且杆轴线通过铰结中心的杆系结构. 若所有杆件均在同一平面内，称为平面桁架，否则为空间桁架. 平面桁架只能承受平面荷载，且作用在结点上(杆自重除外)，杆中仅产生轴力(拉或压)，称为二力杆. 用平衡方程能求出全部内力的桁架结构，称为静定桁架；若必须考虑变形协调条件求解桁架内力的结构，称为超静定桁架. 其解法有力法、位移法、矩阵位移法等. 空间桁架(静定或超静定)若是由几个平面桁架组成的，则可分解成平面桁架来求解，否则必须按空间问题求解. 桁架在工程中应用很广，如屋架、塔架等.

杆系结构(structure of bar system) 一种类型的结构. 指杆件(直杆或曲杆)间用结点连结组成的结构. 各杆在结点处不允许有相对线位移. 若各杆在结点处允许有相对转动，则为铰结点，否则为刚结点. 铰结点不能承受和传递力矩，刚结点则反之. 全部结点均为铰结点的结构为桁架；具有全部刚结点或部分刚结点组成的结构称为刚架或框架. 若所有杆件均在同一平面，则称为平面杆系结构，该结构只能承担平面内的外力，否则称为空间杆系结构. 分析杆系结构时，应首先分清是平面还是空间、是静定还是超静定结构，再根据不同类型结构的特点采用相应的计算方法求解.

位移(displacement) 力学的基本概念之一. 指结构体系各点位置的移动和截面的转动. 其原因是荷载的作用或温度的变化、支座的移动等. 体系中各点或截面沿某方向的移动称为线位移，某两点或两截面的相对移动称为相对线位移，截面沿某方向的转动称为角位移，相对转动称为相对角位移. 梁、板壳结构的挠度是线位移，其截面的转动为角位移. 位移与时间无关的，称静位移，与时间有关的，称动位移. 以上这些位移统称为广义位移(线位移、角位移、相对线位移、相对角位移). 与广义位移相对应的力系称为广义力.

广义位移(generalized displacement) 见"位移".

广义力(generalized force) 力学概念. 指结构位移计算中，求广义位移时需施加的相应力系. 求线(角)位移需加单位力(或力矩)，求相对线(角)位移时需加一对大小相等、方向相反的力(或力矩). 引起结构内力和变形的外来因素也可称为广义力，如集中力、力矩、均布力、超静定结构支座移动或温度改变施加给结构的力.

单位荷载法(unit-load method) 亦称麦克斯韦-莫尔法. 根据虚功原理求结构位移的一种方法：若求结构某点 i(或某两点)由于外荷载、支座移动和温度改变产生的位移(或相对位移)Δ_i，则需在该点求位移的方向时施加虚设的单位广义力 $p_i=1$(求线位移时加力；求转角时加力矩；求相对线位移或角位移时在两点加一对大小相等、方向相反的力或力矩). 由虚功原理，虚设力系在真实位移上所作的虚外功，在数值上等于虚设力系在结构中引起的虚内力在真实应变过程中所作的虚内功，即

$$1\cdot\Delta_i=\sum\int\frac{\overline{N}N_{\mathrm{p}}}{GA}\mathrm{d}s+\sum\int\frac{k\overline{Q}Q_{\mathrm{p}}}{GA}\mathrm{d}s+\sum\int\frac{\overline{M}M_{\mathrm{p}}}{EI}\mathrm{d}s-\sum\overline{R}_cC_c+(\pm)\alpha\sum t_0\omega_{\mathrm{N}}+(\pm)\alpha\sum\frac{\Delta t}{h}\omega_{\mathrm{M}},$$

式中：$\overline{N}$，$\overline{Q}$，$\overline{M}$ 分别为虚设力系引起杆内的轴力、剪力、弯矩；N_{p}，Q_{p}，M_{p} 分别为外荷载引起杆内的轴力、剪力、弯矩；E，G 分别为材料的弹性模量、剪切模量；h 为截面高度；A，I 分别为截面面积、截面惯性矩；α 为材料线膨胀系数；t_0 为截面形心温度改变量；Δt 为上、下边缘温差；$\overline{R}_c$ 为虚设力系引起有支座移动处的相应支反力；C_c 为实际支座移动值；ω_{N}，ω_{M} 分别为虚设力系引起的轴力图形面积、弯矩图形面积，当虚设力系引起杆内力产生的变形趋势与温度引起的杆件变形趋势一致时取正号，否则取负号；$\mathrm{d}s$ 为杆件微长度；$\sum$表示对全部杆件和全部移动支座

求和. 对于桁架只取等式右边第一项、第四项和第五项. 对于梁和刚架，一般取第三、四、五和六项即可. 对于直杆可用图乘法.

单位荷载法是由英国物理学家麦克斯韦(Maxwell, J. C.)于 1864 年、莫尔(Mohr, O.)于 1874 年分别提出的. 该法常用于求解杆系结构、薄壁结构的静定问题和超静定问题位移计算和内力分析.

麦克斯韦-莫尔法(Maxwell-Mohr method)　即"单位荷载法".

力法(force method)　亦称柔度法. 求解超静定结构的一种方法. 是以广义力为基本未知量计算超静定结构的基本方法. 用此法解超静定结构时，是把超静定问题与静定问题加以比较，从中找出由静定过渡到超静定的途径. 为此，先要确定超静定次数，再将多余约束撤去而代之以广义未知力(即基本未知量)，则所得的静定体系称为力法的基本体系，又称基本结构. 该体系是静定的，同时也必须是几何不变的. 对于基本体系，未知力变成了主动力. 为使基本体系的变形与原超静定结构完全相同，必须使基本体系中沿每一多余未知力的方向的位移与原结构中相应的位移相等，此条件称为变形协调条件. 每一个协调条件对应于一个求解多余未知力的补充方程. 例如，对于 n 次超静定结构，即可得到求解 n 个未知力 $x_i(i=1,2,\cdots,n)$的 n 个力法方程：

$$\begin{aligned}&\delta_{11}x_1+\delta_{12}x_2+\cdots+\delta_{1n}x_n+\Delta_{1p}=0,\\&\delta_{21}x_1+\delta_{22}x_2+\cdots+\delta_{2n}x_n+\Delta_{2p}=0,\\&\cdots\cdots\\&\delta_{n1}x_1+\delta_{n2}x_2+\cdots+\delta_{nn}x_n+\Delta_{np}=0,\end{aligned}$$

式中，δ_{ij}为广义力 $x_j=1$ 时引起对应于 x_i 的广义力方向的位移，称柔度系数；Δ_{ip}为外载(p)作用时引起对应于 x_i 的广义力方向的位移. 解此方程即可求得 n 个广义未知力，再利用叠加原理可求出其他内力，如内力 M 的值，$M=M_p+\overline{M}_ix_i$，M_p 为外力引起的弯矩值，$\overline{M}_i$ 为广义力 $x_i=1$ 引起的弯矩值.

柔度法(flexibility method)　即"力法".

柔度系数(flexibility coefficient)　见"力法".

变形协调条件(compatibility condition of deformation)　见"力法".

矩阵力法(matrix force method)　求解超静定结构的一种方法. 是以矩阵为工具，以广义力为基本未知量计算超静定结构的方法.

位移互等定理(reciprocal theorem of displacement)　亦称互等位移定理. 功的互等定理的一种特殊情形. 材料力学、结构力学和弹性力学中，反映不同点上广义位移相互关系的定理. 是由英国物理学家麦克斯韦(Maxwell, J. C.)于 1864 年提出. 该定理可表述为：在任一线性弹性体上，在广义力 p_1 单独作用下，引起与广义力 p_2 相对应的广义位移 Δ_{21}(第一个下标为引起位移的地点和方向，第二个下标为引起位移的原因)，在数值上等于 p_2 单独作用下引起与广义力 p_1 相对应的广义位移 Δ_{12}，即 $\Delta_{21}=\Delta_{12}$. 位移 Δ_{ij}与广义力 p_j 的比值是一常数，记为 δ_{ij}，故该定理也可表示为：在任一线弹性体系中，由广义力 p_1 所引起的与广义力 p_2 相对应的位移影响系数 δ_{21}，等于广义力 p_2 所引起的与广义力 p_1 相对应的位移影响系数 δ_{12}，即 $\delta_{21}=\delta_{12}$. 或者说，由单位广义力 $p_1=1$ 所引起的与广义力 p_2 相对应的位移，在数值上等于由广义力 $p_2=1$ 所引起的与广义力 p_1 相对应的位移.

互等位移定理(reciprocal theorem of displacement)　即"位移互等定理".

互等功定理(reciprocal theorem of work)　功的互等定理的别称.

反力互等定理(reciprocal theorem of reaction)　功的互等定理的一种特殊情形. 是结构力学和弹性力学中反映不同的约束支座反力相互关系的定理. 可表示为：在任一线性弹性体系中，在支座 i 发生广义位移(角位移或线位移)C_i 时，在支座 j 引起的与 j 约束相对应的支座反力系数 $r_{ji}=R_{ji}/C_i$(R_{ji}为在支座 i 发生广义位移 C_i 时，在 j 约束的相应支座反力)，在数值上等于在支座 j 发生广义位移 C_j 时，在支座 i 引起的与 i 约束的相应支座反力系数 $r_{ij}=R_{ij}/C_j$，即 $r_{ji}=r_{ij}$. 或者说，由单位位移 $C_i=1$ 所引起的与位移 C_j 相应的反力，在数值上等于由单位位移 $C_j=1$ 所引起的与位移 C_i 相应的反力.

位移法(displacement method)　亦称刚度法. 结构的一种求解方法. 是以结点广义位移(角位移和线位移)为基本未知量求解结构的一种方法. 用此法计算结构时，其基本作法是先拆散，后组装——即先把整体结构离散成单个构件(或称为单元)进行分析；组装时，首先根据在结点处各杆件的变形要协调一致，其次根据装配好的结点要满足平衡条件二原则进行整体分析. 位移法方程可由基本结构法建立，也可直接利用平衡条件建立. 如用基本结构法解超静定刚架，则先要确定结构全部结点的广义位移 z_i ($i=1,2,\cdots,n$)，其次将广义位移加以相应的约束(对角位移附加刚臂，对线位移附加支杆)，使原结构变成一系列的单跨超静定梁所组成的基本结构. 为使基本结构的约束条件与原结构相同，必须使附加约束上产生的约束 R_i(约束力矩或约束反力)为零. 即基本结构在外荷载作用下，在附加约束上产生的约束力 R_{ip}与分别解除附加约束使之产生相应的位移 z_i，而在广义位移 z_j 的约束上产生的广义力 R_{ij}之

和为零，此条件为位移法的平衡条件. 例如对于 n 个广义位移，即可得到 n 个平衡方程

$$R_{11}+R_{12}+\cdots+R_{1n}+R_{1p}=0,$$
$$R_{21}+R_{22}+\cdots+R_{2n}+R_{2p}=0,$$
$$\cdots\cdots$$
$$R_{n1}+R_{n2}+\cdots+R_{nn}+R_{np}=0.$$

因真实位移 z_i 为未知，故 R_{ij}不能直接求出，需用解除附加约束使之产生相应的单位位移，而在广义位移的约束上产生的广义力 r_{ij}(称刚度系数)，再乘以真实广义位移 z_j，求得广义力 $R_{ij}=r_{ij}\cdot z_j$，即可得到平衡方程组

$$\begin{cases} r_{11}z_1+r_{12}z_2+\cdots+r_{1n}z_n+R_{1p}=0, \\ r_{21}z_1+r_{22}z_2+\cdots+r_{2n}z_n+R_{2p}=0, \\ \cdots\cdots \\ r_{n1}z_1+r_{n2}z_2+\cdots+r_{nn}z_n+R_{np}=0. \end{cases}$$

解方程组得出 n 个广义位移 z_i，然后利用叠加法可得出结构各杆内力. 如各杆弯矩值 M 为

$$M=M_p+\overline{M}_i\cdot z_i,$$

其中 M_p 为外载荷作用在基本结构上各杆的弯矩值，$\overline{M}_i$ 为 $z_i=1$ 时产生的弯矩值.

刚度法(stiffness method)　即"位移法".

刚度系数(stiffness coefficient)　见"位移法".

刚度(stiffness)　结构力学的基本概念之一. 是受外力作用的构件或结构抵抗变形的能力. 结构的刚度除取决于组成结构的材料弹性模量、剪切模量之外，还取决于结构的几何形状、约束条件以及外力的作用形式. 对于一些需严格限制变形的结构(如机翼、高精度的装配件等)，必须通过刚度分析来控制变形. 许多建筑结构、机械设备也要通过刚度分析来防止过大的变形和发生振动、颤振、共振或失稳，故在工程设计中研究结构刚度是一项重要内容. 在结构力学位移法分析中，为确定结构的变形和内力，也要分析各部分的刚度. 与刚度相对应的是柔度概念，它表示结构在一定外力作用下变形的大小. 在数值上柔度和刚度互为倒数关系. 柔度这一概念常用于力法分析中.

柔度(flexibility)　见"刚度".

结构矩阵分析(matrix structural analysis)　一种对结构的分析方法. 是以传统结构力学为理论基础，以矩阵为数学表述形式，以电子计算机为计算手段，三位一体进行的结构分析. 其步骤为：

1. 把结构离散成若干单元(在杆系结构中，一般把每根杆件取作一个单元)，进行单元分析，得出单元的刚度方程或柔度方程.

2. 将这些在结点处满足变形协调条件和平衡条件的单元集合成整体结构，得出整体结构的基本方程.

3. 解基本方程，求出结点位移或结构内力.

结构矩阵分析的基本方法是矩阵位移法和矩阵力法.

矩阵位移法(matrix displacement method)　一种对结构的分析方法. 指以矩阵为数学表达形式，以电子计算机作为计算手段，并以结构中的结点广义位移(线位移和角位移)为基本未知量分析结构的方法. 在结构力学中，该方法是有限单元法的雏形，因此，也称杆系结构的有限单元法. 用此法时，先将结构分解为若干个单元(一般一根杆件为一单元)进行单元分析，建立单元刚度方程

$$[k]^e\{\Delta\}^e=\{F\}^e,$$

式中 $\{F\}^e$，$[k]^e$ 和 $\{\Delta\}^e$ 分别为在局部坐标系中单元杆端力向量、单元刚度矩阵和单元杆端位移向量. 其次，进行整体分析，根据结点平衡条件和变形协调条件，将局部坐标系中的单元杆端力、位移和单刚转换到整体坐标系上，集合成整体平衡方程

$$[k]\{\Delta\}=\{p\},$$

式中 $\{p\}$，$\{\Delta\}$ 分别为结点载荷、位移，$[k]$ 为结构刚度矩阵. 由整体平衡方程解出结点位移 $\{\Delta\}$，代入单元刚度方程，即可得出各杆内力，给出内力图.

转角位移法(slope-deflection method)　一种对结构的求解方法. 指以结点位移(角位移和线位移)为未知量求解连梁和刚架等结构的一种方法. 是位移法的一种，该法由本迪克森(Bendixen, A.)于1914年提出. 用该法时，首先研究每一杆件在外载和杆端发生位移(角位移和两杆端的相对线位移)情况下引起的杆端力，利用叠加原理得出杆端力与转角、线位移、外载之间关系式. 如对于 AB 杆，其关系式为

$$M_{AB}=2i\left(2\varphi_A+\varphi_B-3\frac{\Delta_{AB}}{l}\right)+M_{AB}^F,$$
$$M_{BA}=2i\left(\varphi_A+2\varphi_B-3\frac{\Delta_{AB}}{l}\right)+M_{BA}^F,$$
$$Q_{AB}=-\frac{6i}{l}\left(\varphi_A+\varphi_B-2\frac{\Delta_{AB}}{l}\right)+Q_{AB}^F,$$

式中 M_{AB}，M_{BA} 和 Q_{AB} 分别为 AB 杆两端杆端力矩和 A 端剪力；M_{AB}^F，M_{BA}^F 和 Q_{AB}^F 分别为外载作用下产生的 AB 杆两端杆端力矩和 A 杆端剪力；φ_A，φ_B 和 Δ_{AB} 分别为 A，B 两杆端未知转角和两杆端未知相对线位移；$i=EI/l$ 为线刚度，E 为材料弹性模量，I 为杆的截面惯性矩；l 为 AB 杆杆长. 上式写成矩阵形式为

$$\begin{Bmatrix} M_{AB} \\ M_{BA} \\ Q_{AB} \end{Bmatrix}=\begin{bmatrix} 4i & 2i & -\frac{6i}{l} \\ 2i & 4i & -\frac{6i}{l} \\ -\frac{6i}{l} & -\frac{6i}{l} & \frac{12i}{l^2} \end{bmatrix}\begin{Bmatrix} \varphi_A \\ \varphi_B \\ \Delta_{AB} \end{Bmatrix}+\begin{Bmatrix} M_{AB}^F \\ M_{BA}^F \\ Q_{AB}^F \end{Bmatrix},$$

式中 3×3 矩阵称为刚度矩阵，其系数为刚度系数．其次取结点或部分杆件为隔离体，由平衡条件（结点平衡及截面平衡）得出包括各广义位移为未知量的代数方程组，并由此解出位移，再回代到转角位移方程，求出各杆端内力．

力矩分配法（moment distribution method） 亦称克劳斯法．一种求结构杆端力矩的方法．是以位移为基础的一种逐次渐近地求结构各杆端（与刚结点连结的杆端）力矩的方法．该法由克劳斯（Cross，H.）于 1930 年首先提出．该法主要用于计算连梁和无侧移刚架．用该法时，先将可能转动的刚结点用假想附加刚臂（只阻止转动而不阻止移动）固定起来，加载后，刚臂中就产生附加力矩（由杆的固端力矩算得），然后逐个放松附加刚臂，消除刚臂中的附加力矩，使之恢复结构的真实变形状态．如若首先放松结点 i 的附加刚臂，则刚结于 i 结点的所有各杆就会变形，附加力矩随即按一定比例分配于连结 i 结点的各杆近端，称分配力矩，而各杆远端也将引起一定力矩，称传递力矩．然后，再将 i 结点重新加上附加刚臂，在消除于结点 i 相邻的结点 j 刚臂中所附加的不平衡力矩时，ij 杆的 i 杆端将重新得到 j 杆端传来的传递力矩，仍由结点 i 的附加刚臂来承担．再次放松结点 i 的附加刚臂，使之消除不平衡力矩．如此循环进行，直到各结点上的附加刚臂中附加力矩都趋于零为止．循环计算完后，将各杆端固端力矩和各次分配力矩、传递力矩相加，得各杆端的实际力矩值．

分配系数（distribute coefficient） 进行力矩分配时所用的一个数值．指刚结于某结点处的某杆的转动刚度与交于该结点各杆的转动刚度之和的比值．例如，交于 A 结点的 A_j 杆的分配系数（j 为 A_j 杆远端编号）为

$$\mu_{A_j}=\frac{S_{A_j}}{\sum_i S_{A_i}},$$

式中 μ_{A_j} 和 S_{A_j} 分别为 A_j 杆 A 杆端的分配系数和转动刚度，$\sum_i S_{A_i}$ 为交于 A 结点所有各杆的 A 杆端的转动刚度之和．

传递系数（transfer coefficient） 进行力矩分配时所用的一个数值．指当某杆近端有转动时，远端弯矩与近端弯矩的比值．如 AB 杆近端弯矩为 M_{AB}，远端弯矩为 M_{BA}，则传递系数

$$c_{AB}=\frac{M_{BA}}{M_{AB}}.$$

对于等截面直杆，传递系数 c 随远端的支承情况而异，如远端固定，$c=1/2$；远端滑动支承，$c=-1$；远端为铰支，$c=0$.

转动刚度（rotational stiffness） 结构力学的基本概念之一．指杆端对转动的抵抗能力．它在数值上等于使杆端产生单位转角时需要施加的力矩值．对于等截面直杆，转动刚度 S 将随远端支承情况而异．如远端为固定支座，$S=4i$，其中 $i=EI/l$ 为线刚度，E 为杆件的弹性模量，I 为截面惯性矩，l 为杆件长度．远端为简支时，$S=3i$．远端为滑动支承时，$S=i$．远端为自由端时，$S=0$．

单元刚度矩阵（element stiffness matrix） 亦称转换矩阵，或简称单刚．用矩阵形式表示的一种单元内部的关系式．指在杆系结构中，单元杆端力与杆端位移之间的转换关系式．在局部坐标系中，单元 e 的刚度方程为

$$\{F\}^e=[k]^e\{\delta\}^e,$$

其中：e 表示单元号；$\{F\}^e$ 为 e 单元杆端力向量；$\{\delta\}^e$ 为杆端位移向量；$[k]^e$ 为单元刚度矩阵．对于平面问题自由式单元，有六个杆端力和六个杆端位移．单刚时的 $[k]^e$ 为 6×6 阶矩阵，它具有对称性、奇异性，它的每一个元素称为单元刚度系数．如对于平面问题梁式单元，单元 e 杆端的编码为 i（左杆端）和 j（右杆端），当 i 杆端发生单位转角（$\theta_i=1$）时，引起 i，j 两杆端的弯矩值分别为

$$M_i^e=4i^e,\quad M_j^e=2i^e,$$

式中 $i=EI/l$ 为线刚度，E 为材料弹性模量，I 为主惯性矩，EI 为抵抗弯曲的能力，l 为梁跨度．当 e 单元两端同时发生转角为 θ_i^e 和 θ_j^e 时，引起两杆端弯矩值为

$$M_i^e=4i^e\theta_i^e+2i^e\theta_j^e,$$
$$M_j^e=2i^e\theta_i^e+4i^e\theta_j^e.$$

写成矩阵形式为

$$\begin{pmatrix}M_i\\M_j\end{pmatrix}^e=\begin{bmatrix}4i^e & 2i^e\\2i^e & 4i^e\end{bmatrix}\begin{pmatrix}\theta_i\\\theta_j\end{pmatrix}^e=[k]^e\{\theta\}^e,$$

其中

$$[k]^e=\begin{bmatrix}k_{ii} & k_{ij}\\k_{ji} & k_{jj}\end{bmatrix}=\begin{bmatrix}4i^e & 2i^e\\2i^e & 4i^e\end{bmatrix}$$

为单刚．

转换矩阵（conversion matrix） 即“单元刚度矩阵”．

整体刚度矩阵（global stiffness matrix） 亦称总刚度矩阵，或简称总刚．表征结构整体性质的矩阵．是杆系结构中，结点力与结点位移之间的转换矩阵．整体结构的刚度方程为 $\{F\}=[k]\{\Delta\}$，其中 $\{F\}$ 为结点力向量，$\{\Delta\}$ 为结点位移向量，$[k]$ 为整体刚度矩阵．总刚中的某一元素 k_{ij} 称为刚度系数，它表示当第 j 个结点位移分量 $\delta_j=1$（其他各结点位移分量全为零）时，所产生的第 i 个结点位移方向所施加的结点力 F_i．总刚具有对称性、稀疏性和可逆性．建立总刚的方法有传统位移法和单刚集成法（或称为

直接刚度法).单刚集成法便于实现计算过程的程序化.

单元柔度矩阵(element flexibility matrix) 用矩阵形式表示的一种单元内部的关系式.指在杆系结构中,单元杆端位移用杆端力表达时的联系矩阵.在局部坐标系中,由单元 e 杆端力求杆端位移的柔度方程为

$$\{\delta\}^e=[c]^e\{F\}^e,$$

其中 $\{\delta\}^e=[u_1^e\ u_2^e\ u_3^e\ u_4^e\ u_5^e\ u_6^e]^{\mathrm{T}}$ 为局部坐标系中两杆端的轴向、横向和转角位移;$\{F\}^e=[F_1^e\ F_2^e\ F_3^e\ F_4^e\ F_5^e\ F_6^e]^{\mathrm{T}}$ 为与杆端位移相对应的杆端力;$[c]^e$ 为柔度矩阵.矩阵中每一元素为柔度系数,如 c_{ij} 表示在局部坐标系中第 j 个杆端力 $F_j=1$ 时所产生的第 i 个杆端位移分量 u_i.单元柔度矩阵具有对称性、可逆性.如具有轴向力的梁式单元(局部坐标系与整体坐标系一致时),其杆端位移与杆端力之间的关系式为

$$\begin{Bmatrix} u_3^e \\ u_4^e \\ u_6^e \end{Bmatrix}=\begin{bmatrix} \dfrac{l}{3EI} & 0 & -\dfrac{l}{6EI} \\ 0 & \dfrac{l}{EA} & 0 \\ -\dfrac{l}{6EI} & 0 & \dfrac{l}{3EI} \end{bmatrix}\begin{Bmatrix} F_3^e \\ F_4^e \\ F_6^e \end{Bmatrix},$$

柔度矩阵为

$$[c]^e=\begin{bmatrix} \dfrac{l}{3EI} & 0 & -\dfrac{l}{6EI} \\ 0 & \dfrac{l}{EA} & 0 \\ -\dfrac{l}{6EI} & 0 & \dfrac{l}{3EI} \end{bmatrix}.$$

其中 E 为材料弹性模量;A 为梁元截面面积;I 为截面惯性矩.

整体柔度矩阵(global flexibility matrix) 亦称总柔度矩阵.表征结构整体性质的矩阵.是在杆系结构中,结点位移用结点力表达时的联系矩阵,整体结构有 n 个结点位移的柔度方程为

$$\{\delta\}_{n\times1}=[c]_{n\times n}\{F\}_{n\times1},$$

其中 $\{\delta\}_{n\times1}$ 为结点位移向量;$\{F\}_{n\times1}$ 为结点力向量;$[c]_{n\times n}$ 为整体柔度矩阵,矩阵中 c_{ij} 为柔度系数,它具有对称性、稀疏性和可逆性.建立柔度矩阵的方法有传统力法和单元柔度矩阵集成法(也称为直接柔度法).

结构塑性设计(plastic design of structure) 亦称极限设计.是以结构塑性分析为基础的设计.结构的弹性设计是以应力为依据.而事实上,结构中虽然某一局部已达弹性极限值,但结构的塑性设计考虑到材料的塑性使应力重新分布,整体结构仍能继续承受荷载而不破坏,故此种设计具有能充分利用材料的塑性、简单、合理、并能节省材料等特点.结构的塑性设计先要确定极限荷载 p_j,然后将极限荷载除以安全系数 k 得出容许荷载$[p]$,并以此为依据来进行设计.结构上所能承受的实际荷载 p 必须满足下面强度条件

$$p\leqslant[p]=\frac{p_i}{k}.$$

安全系数 k 是从整体结构所能承受的荷载来考虑的,所以它比弹性设计中的安全系数 k' 更能正确地反映结构的强度储备.在实际问题中,$k>k'$.

极限设计(limit design) 即"结构塑性设计".

结构极限设计(limit design of structure) 结构塑性设计的别称.

极限荷载(limit load) 结构力学的基本概念之一.指由理想弹塑性材料制成的结构丧失承载能力时所能承担的荷载的极限值.这时的结构变成了机构,变形可以任意增大,承载力无法再增加,这种状态为结构的极限状态.极限荷载可根据极限状态的弯矩图由平衡条件推出,也可根据破坏机构的虚功方程求出.如一超静定梁(图1),极限弯矩为 M_j,破坏机构见(图2),由虚功方程

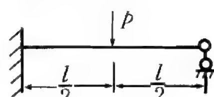


图 1

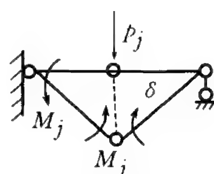


图 2

$$p_j\delta-\frac{6\delta M_j}{l}=0,$$

得极限荷载

$$p_j=\frac{6M_j}{l},$$

其中 δ 为跨中虚位移.结构极限荷载计算特点为:

1. 只需考虑最后的破坏机构,不需考虑结构弹塑性变形的发展过程.

2. 只需考虑静力平衡条件,不需考虑变形协调条件.

3. 不受温度变化、支座移动等因素的影响.

极限弯矩(limit bending moment) 结构力学的基本概念之一.指理想弹塑性材料制成的梁或板截面达到塑性状态时所能承受的最大弯矩.极限弯矩 M_j 的算式为

$$M_j=\iint\limits_{A_1}\sigma_s\eta\mathrm{d}A+\iint\limits_{A_2}\sigma_s\eta\mathrm{d}A,$$

式中 A_1,A_2 分别为受拉区和受压区的面积;σ_s 为材料屈服极限;η 为任意微面积至中性轴的距离.根据整个截面对中性轴静矩为零的条件,有

$$\iint\limits_{A_1}\eta\mathrm{d}A=-\iint\limits_{A_2}\eta\mathrm{d}A=s_1(\eta),$$

其中 $s_1(\eta)$ 为截面 A_1 对中性轴的静矩.这时,极限弯矩的算式为

$$M_j=\sigma_s(s_1+s_2).$$

如矩形截面

$$M_j=\frac{bh^2\sigma_s}{4},$$

其中 b,h 分别为截面的宽度和高度.

塑性铰(plastic hinge)　结构力学的基本概念之一.指由理想弹塑性材料制成的梁或刚架,当某截面达到塑性流动阶段时,在极限弯矩保持不变的情况下,两个无限靠近的相邻截面可产生有限的相对转动的断面.该处的曲率变化率可以任意增大并出现间断,这种情况与带铰的截面相似.因此,当截面弯矩达到极限弯矩时,这种截面称为塑性铰.它与普通铰的区别是:

1.普通铰不传递弯矩.而它能传递极限弯矩.

2.普通铰为双向铰.而它是单向铰,即只能沿弯矩增大方向发生有限的相对转动;如果沿反方向变形,则该截面立即恢复其弹性刚度,而不再具有铰的性质.

塑性铰一般出现在弯矩较大的地方,如集中力作用处、固定支座处和分布载荷作用下剪力为零处.当结构中形成足够多的塑性铰后,变成机构,丧失承载能力.塑性铰的概念在梁和刚架的极限分析中已广泛应用.

塑性铰线(plastic hinge line)　结构力学的基本概念之一.指由理想弹塑性材料制成的板或壳体结构,其中某截面承受的弯矩值达到极限弯矩值,该截面达到塑性流动阶段时所形成的塑性铰的连线.它的性质与塑性铰相同.当板、壳结构产生足够多的塑性铰线后,能使其结构变成机构,处于极限状态.在板、壳极限分析中,应用虚功原理可求得极限荷载值,一般都是极限荷载的上限.塑性铰线在矩形板、多边形板、圆板、旋转壳等的极限分析中已获得广泛应用.

结构极限分析(limit analysis of structure)　研究结构在塑性状态下各种特性的方法.由理想弹塑性材料制成的结构,当所承受的荷载增大到极限值时,在结构中形成了足够多的塑性铰或塑性铰线后,结构变成了机构,从而丧失承载能力,在不再加大荷载的情况下,变形续继增大,此种状态称为极限状态,所对应的荷载为极限荷载.通过极限分析,可求出极限荷载值、此状态下的应力分布规律及破坏形式.此理论在连续梁、刚架、桁架、组合结构、矩形板、圆板、多边形板、旋转壳等结构中已得到广泛应用.极限分析除能找出该结构所能承受的最大荷载值外,还能找出合理的结构形式.

上限定理(upper bound theorem)　亦称极小定理.结构分析理论中的一个重要定理:对于比例加载,由理想刚塑性材料制成的结构,按照各种机动许可的单向破坏机构(沿荷载方向作单向运动)中相对应的可破坏荷载(p_k),是极限荷载的上限值,或者说极限荷载是可破坏荷载中的极小者.利用外力虚功与内力虚功相等的原则,便可求出该机构所对应的外荷载.用这种破坏机构计算的荷载值将大于或者等于真实极限荷载值(p_j),即 $p_j\leqslant p_k$.破坏机构通常可通过试验方法得到,故所计算的荷载上限值同真实的结构极限值十分接近.上限定理在结构极限分析中已得到广泛应用.

极小定理(minimum theorem)　即"上限定理".

下限定理(lower bound theorem)　亦称极大定理.结构极限分析理论中的一个重要定理:对于比例加载并由理想刚塑性材料制成的结构,按照各种静力可能而又安全的内力分布状态所求得的可接受荷载(p_s),将小于或者等于真实极限荷载(p_j),即 $p_s\leqslant p_j$.可接受荷载的最大值就是极限荷载(p_j)的下限值,极限荷载是可接受荷载中的极大者.

极大定理(maximum theorem)　即"下限定理".

结构稳定性(stability of structure)　结构力学的基本概念之一.指研究弹性结构平衡状态能否保持稳定的理论.结构平衡状态有三种形式:稳定平衡、随遇平衡(或称中性平衡)和不稳定平衡.设结构原来处于某个平衡状态,此后若有微小扰动,使之稍微偏离其原来位置,在扰动消失后,如果能回到或有趋势回到原平衡位置,则称原平衡状态为稳定平衡;如若结构继续偏离,不能回到原来位置,则称不稳定平衡状态.这时结构丧失稳定性,简称失稳,又称丧失第一类稳定性.除此情形外,还有丧失第二类稳定性,其特征是平衡形式并不发生质变,而是结构丧失承载能力,当荷载达临界荷载时,即使不继续加大荷载,甚至减少荷载,变形也会自行增加.不论是丧失第一类或第二类稳定性,在工程中都是不容许的.随遇平衡通常是从稳定平衡状态向不稳定平衡状态过渡的中间状态,也可归入不稳定平衡的范畴.随所受荷载的增大,结构由稳定平衡状态经随遇平衡状态过渡到不稳定平衡状态,相应于结构的随遇平衡状态的最小荷载称为临界荷载.在工程结构的稳定性研究中,应求出结构的临界荷载,并保证其工作外荷载小于临界荷载.

临界荷载(critical load)　见"结构稳定性".

结构动力学

结构动力学(structural dynamics)　结构力学的一个分支.它研究结构在动荷载(周期荷载、冲击荷载及随机荷载等)作用下的响应(位移、应力、应变等随时间变化的规律)、承载能力、动力特性(如自由

振动频率)等.结构动力分析与静力分析的区别在于是否考虑运动加速度.根据达朗贝尔原理,可把动力问题转化为静力问题来处理,考虑结构因振动产生惯性力条件下的动平衡问题,同时还需考虑阻尼以及因变形而产生的弹性力.研究动力特性还是为了避免共振现象的发生.研究方法主要包括理论分析和实验研究.理论分析首先要建立含空间和时间变量的微分方程(线性或非线性微分方程),结合边界条件,用解析方法或用数值方法求解.实验研究分实物动荷载实验和模型动荷载实验,进行各种参数(如振幅、频率、相位、应力、变形等)的测量和计算.

结构固有振动(natural vibration of structures) 亦称主振动.结构的自由振动.指无阻尼线弹性结构体系在无外加激励条件下的单频率振动.在结构固有振动中,各质点 $M_i(i=1,n,\cdots,\infty)$ 以相同频率作简谐运动,且彼此同相或反相.线弹性结构为无限自由度体系,有无限个固有振动,若离散成 n 个自由度体系,则有 n 个固有振动.一般地,结构的自由振动是 n 个固有振动的线性叠加.

主振动(principal vibration) 即"结构固有振动".

结构的固有频率(natural frequency of structure) 结构振动的特性.指线弹性结构主振动频率.无阻尼结构作主振动时,结构各质点以相同频率作简谐运动.结构为无限自由度体系,有无限个频率.若将结构离散为 n 个自由度体系,将有 n 个频率 $f_{ni}(i=1,2,\cdots,n)$(或角频率 $\omega_{ni}=2\pi f_{ni}$).这些频率与振动的初始条件无关,仅取决于结构的惯性和弹性参数,故称固有频率.固有频率与固有模态是线弹性结构固有振动的特性.通常将固有频率按由小到大的顺序排列:

$$f_{n1}\leqslant f_{n2}\leqslant\cdots\leqslant f_{nm},$$
$$\omega_{n1}\leqslant\omega_{n2}\leqslant\cdots\leqslant\omega_{nm},$$

依次称为第一(一阶)频率(或基本频率)、第二(二阶)频率等.结构受到周期激励时,常需计算结构的最低若干个固有频率,以检查有无共振危险.固有频率的常用单位是赫兹(Hz),固有圆频率的常用单位是弧度/秒(rad/s).

结构自振周期(free vibration period of structure) 结构振动的特性.指线弹性结构自由振动时,对其主振型结构体系重复振动一次所需最短时间,记为 T(单位为秒).结构体系为无限自由度体系,有无限个振动周期($T_i\,|\,i=1,2,\cdots,\infty$),若将结构离散为 n 个自由度体系,将有 n 个周期 $T_{ni}(i=1,2,\cdots,n)$,第 i 个振型的自振周期的倒数($1/T_{ni}$)称结构第 i 个振型的自振频率,记为 f_{ni},它表示第 i 个振型单位时间内振动的次数,常用单位是 1/秒 (s^{-1}),称为赫兹(Hz).第 i 个振型在 2π 单位时间内振动的次数,称结构第 i 个振型的圆频率,记为 ω_{ni},即 $\omega_{ni}=2\pi f_{ni}$.通常将周期按由大到小(或由长到短)顺序排列:

$$T_{n1}\geqslant T_{n2}\geqslant\cdots\geqslant T_{nn},$$

依次称为第一周期(或基本周期,或长周期)、第二周期等.结构在周期载荷作用下,产生周期性振动,在随机载荷作用下,将产生非周期性振动.

结构自振频率(free frequency of structure) 见"结构自振周期".

结构圆频率(circular frequency of structure) 见"结构自振周期".

结构振型(vibration mode of structure) 结构振动术语.指线弹性结构各点以特定的频率作简谐振动时驻波的形态.实际上,也就是弹性结构的实模态.

动力放大系数(dynamic amplification coefficient) 一个数值.是单自由度体系受迫振动中其动位移(y_p)与静位移(y_{st})之比 $\eta=y_p/y_{st}$.该系数不仅与干扰力频率 θ 和体系圆频率 ω 之比 θ/ω 有关,而且与阻尼比 ξ 有关,即

$$\eta=\left[\left(1-\frac{\theta^2}{\omega^2}\right)^2+4\xi^2\frac{\theta^2}{\omega^2}\right]^{-\frac{1}{2}},$$

其中 $\xi=c/2m\omega$,c 为阻尼系数,m 为质量.当 $\theta\ll\omega$ 时,$\eta\approx1$,动力位移与干扰力 $R(t)$ 同步,可直接将干扰力 $R_0\sin\theta t$ 作为静载 R_0 来计算位移.当 $\theta\gg\omega$ 时,$\eta\approx0$,质量 m 接近于不动或只作微小振动.当 θ 接近于 ω 时,η 增加很快,而且这时阻尼比 ξ 对 η 值有极大的影响.在 $0.75<\eta<1.25$ 的范围内(称共振区),阻尼力大大地减小受迫振动位移,特别是在共振峰附近.η 的最大值并不是发生在 $\theta=\omega$ 处,而是发生在

$$\frac{\theta}{\omega}=\sqrt{1-2\xi^2}$$

处.又因 ξ 很小,故可近似将 $\theta=\omega$ 时的放大系数的值作为最大值,称此时的振动为共振,其动力放大系数为 $1/2\xi$.

梁的振动(vibration of beam) 一种特殊的振动.指直杆的横向弯曲振动,又称为无限自由度体系的振动.对于等截面梁弯曲时的自由振动,微分方程为

$$\frac{\partial^2 w}{\partial t^2}+\omega^2\frac{\partial^4 w}{\partial x^4}=0,$$

式中 w 为挠度,它是坐标 x 和时间 t 的函数,$\omega=(EI/\overline{m})^{1/2}$ 为圆频率,E 为材料弹性模量,I 为梁截面惯性矩,$\overline{m}$ 为等截面梁的均布质量.用分离变量法可得微分方程的全解

$$w(x,t)=\sum_{n=1}^{\infty}a_nY_n(x)\sin(\omega_nt+\alpha_n),$$

式中 a_n,α_n 为待定常数,由梁的初始条件确定;ω_n($n=1,2,\cdots$)为无限个自振频率;$Y_n(x)$为第 n 个主振型,

$$Y_n(x)=c_1\text{ch}\lambda_nx+c_2\text{sh}\lambda_nx+c_3\cos\lambda_nx+c_4\sin\lambda_nx$$

$$\left(\lambda_n=\sqrt[4]{\frac{\omega_n^2\overline{m}}{EI}}\right),$$

对应每个频率,由梁的边界条件可求出 $c_1\sim c_4$ 的一组比值,得到相应的主振型. 用瑞利-里茨法,可近似求前几个频率,也可用集中质量法求最低频率或较高次频率.

杆的剪切振动(shear vibration of bar)　一种特殊的振动. 是直杆只考虑剪切变形的振动. 对高宽比小于 3 的多层建筑,可简化为剪切型悬臂杆的振动. 等截面直杆剪切变形时的自由振动微分方程为

$$\frac{\partial^2w}{\partial x^2}-\frac{\overline{m}}{K}\frac{\partial^2w}{\partial t^2}=0,$$

式中 w 为垂直杆纵轴的位移;$\overline{m}$ 为等截面杆的均布质量;$K=GA/k$ 为剪切刚度,G 为剪切模量,A 为横截面面积,k 为截面修正系数. 用分离变量法可得微分方程的全解

$$w(x,t)=\sum_{n=1}^{\infty}Y_n(x)\sin(w_nt+\alpha_n),$$

式中 α_n 为待定系数,由初始条件确定;$\omega_n=\lambda_n\cdot\sqrt{K/\overline{m}}$为无限个自振频率,$\lambda_n$ 为与频率有关的特征值;$Y_n(x)$ 为第 n 个主振型,表达式为

$$Y_n(x)=c_1\sin\lambda_nx+c_2\cos\lambda_nx.$$

对应每个频率,由剪切杆的边界条件可求出 c_1,c_2 的一组比值,确定 λ 的特征方程,可求出 ω_n 及相应的主振型 $Y_n(x)$. 剪切自振频率只与材料剪切模量成正比,而与截面尺寸无关 .

结构线性振动(linear vibration of structure)　用线性系统来描述的结构振动. 线弹性结构受到与位移成正比的恢复力和与速度成正比的阻尼力作用的振动. 线弹性结构为无限自由度体系的线性振动,可离散成单自由度或多自由度(n 个)体系的线性振动. 对多自由度体系的受迫振动,其微分方程为

$$[M]\{\ddot{u}\}+[C]\{\dot{u}\}+[K]\{u\}=\{R(t)\},$$

式中$[M]$,$[C]$,$[K]$分别为常系数质量矩阵、阻尼矩阵和刚度矩阵,均为 $n\times n$ 阶;$\{u\}$,$\{\dot{u}\}$和$\{\ddot{u}\}$分别为各质点位移、速度和加速度列向量;$\{R(t)\}$为干扰力列阵,均为 $n\times1$ 阶. 一般地,上式中三个矩阵并不都是对角阵,方程组是耦合的,采用坐标变换手段,使方程组解耦,变成 n 个独立方程,可用杜哈梅积分求解. 这种方法称为正则坐标分析法,或主振型分解法,或主振型叠加法. 另外还可用数值积分法求解动力响应问题. 若阻尼矩阵$[C]$为零,则为无阻尼振动;若无干扰力的作用,则为线弹性结构的自由振动.

结构非线性振动(nonlinear vibration of sturcture)　用线性系统来描述的结构振动. 弹塑性结构振动时,恢复力与位移不是线性的,或阻尼力与速度也不是线性比例的结构振动,称为结构非线性振动. 弹塑性结构为无限自由度体系的振动,其振动微分方程是非线性的(有物理非线性、几何非线性或混合非线性),可将无限自由度体系离散成单自由度或多自由度(n 个)体系的非线性振动,其振动微分方程为

$$[M]\{\ddot{u}\}+[C]\{\dot{u}\}+[K]\{u\}=\{R(t)\},$$

式中$\{u\}$,$\{\dot{u}\}$和$\{\ddot{u}\}$分别为各质点位移、速度和加速度向量;$[M]$,$[C]$,$[K]$分别为 n 阶常系数质量矩阵、阻尼矩阵和刚度矩阵,它们均为随时间而变化的矩阵. 该方程不能用振型叠加原理求解,其动态响应常用数值积分求解.

直接数值积分法(direct numerical integration)　一种求解体系动力响应的方法. 是直接完成对动力微分方程的逐步积分得到体系动力响应的方法. 如 n 个自由度体系的动力方程为

$$[M]\{\ddot{u}\}+[C]\{\dot{u}\}+[K]\{u\}=\{p(t)\},$$

式中$[M]$,$[C]$,$[K]$分别为 n 阶质量矩阵、阻尼矩阵和刚度矩阵;$\{u\}$,$\{\dot{u}\}$和$\{\ddot{u}\}$分别为质点位移、速度和加速度向量;$\{p(t)\}$为作用在体系各质量上的强迫力向量. 求解方程的动力响应,可用振型叠加法、有限差分法和数值积分法. 在一般情况下,动力方程中三个矩阵并不都是对角阵. 对于方程组是耦合的情况,采用坐标变换,可使方程组解耦,变成 n 个独立方程,

$$\{\ddot{x}\}+[\Delta]\{\dot{x}\}+[\Omega^2]\{x\}=[\varphi]^{\mathrm{T}}\{R(t)\},$$

其中$[\Delta]$ 为对角阵,元素值为 $2\omega_i\xi_i$,ξ_i 是第 i 振型阻尼比,$[\varphi]$ 是关于$[M]$ 阵规格化的特征向量 φ_1,φ_2,$\cdots$,$[\Omega^2]$ 是由特征值($\omega_1^2,\omega_2^2,\cdots,\omega_n^2$) 组成的对角阵,$\{x\}$为新坐标系的正则坐标,$\{R(t)\}$为广义荷载. 对于不耦合的方程组,可用杜哈梅积分精确求解.

直接数值积分法,不必首先说明在特征向量上的平衡关系,而是直接地完成对动力微分方程的逐步积分. 从 20 世纪 50 年代,侯伯特(Houbolt,J.C.)提出积分格式开始至今,众多科学工作者提出了多种积分格式,如多步格式、两步格式和单步格式. 从表达形式看,可分为显式和隐式格式,显式格式一般为有条件稳定性格式,隐式格式为无条件稳定格式. 1973 年,威尔森(Wilson,E.L.)提出研究积分格式稳定性问题的基本方法,是用格式中的$[A]$矩阵的特征值 λ_i 来判断其稳定性,即$[A]$矩阵的谱半径 $\rho(A)=\max|\lambda_i|\leqslant1$ 为稳定准则. 积分格式的好坏,

除无条件稳定之外，还要考虑人工阻尼(控制数值逸散)问题、振幅衰减以及周期增长和二阶精度等问题.

风荷载(wind loading)　结构力学的基本概念之一. 指风对结构(如房屋、桥梁、烟囱、电视塔、输电塔等高耸结构物)的作用力、力矩以及表面分布压力. 真实的瞬时风荷载可以分为平均风荷载和脉动风荷载. 脉动风荷载是随受载面的位置、风速和时间等变化的随机荷载，由它引起的结构物的动力响应，称为风振. 响应的计算应采用多自由度体系或无限自由度体系，并按随机振动理论，用振型分解法分析各种结构. 中国在风荷载规范中用

$$\omega_k = \beta_z \cdot \mu_s \cdot \mu_z \cdot \omega_0$$

计算垂直于结构物表面上的风荷载标准值 ω_k. 式中 ω_0 为基本风压(kN/m^2)，它取决于风速(v)以及空气的密度(ρ)；β_z 为结构物离地面高度为 z 处的风振系数，它反映了结构的动力效应，即考虑在随机荷载作用下的各种振型的影响；μ_s 为体型系数，它是作用在结构物表面所引起的实际压力(或吸力)与来风压的比值，主要与结构的体型和尺度有关；μ_z 为风压高度变化系数，它反映在大气边界层内，当气压场随高度不变时，风速随高度增加的规律，主要取决于地面粗糙度和温度垂直梯度.

风振系数(coefficient of wind vibration)　结构动力学的基本概念之一. 是脉动风压对结构的不利影响. 对于刚度小的高耸结构物和高层建筑，必须考虑脉动风压引起的结构的动力效应. 对于多自由度体系或无限自由度体系，通过随机振动理论，采用振型分解法求解. 外形和质量沿高度不变的等截面结构，一般可只考虑第一振型的影响，按无限自由度体系导出沿高度 z 处的风振系数(β_z). 在中国的荷载规范中，按下式计算：

$$\beta_z = 1 + \frac{\xi \gamma \varphi_z}{\mu_z},$$

其中 ξ 为脉动增大系数，其近似计算公式为

$$\xi = \sqrt{1 + x^2 \frac{\pi}{6\zeta} \Big/ (1 + x^2)^{4/3}}$$

$$(x = 30/(\omega_0 T_1^2)^{1/2})$$

是基本风压(ω_0)与结构自振周期(T_1)有关的系数；ζ 为结构的阻尼比；φ_z 为结构的振型系数，当高耸结构按弯曲型考虑时，近似取为

$$\frac{(6z^2H^2 - 4z^3H + z^4)}{3H^4},$$

按剪切型考虑时，近似取为

$$\tan\left[\frac{\pi}{4}\left(\frac{z}{H}\right)^{0.7}\right],$$

其中 H 为结构物高度；γ 为考虑风压脉动及其空间脉动的影响系数，可根据总高度 H 及其与迎面宽度的比值，从给出的规范表中查出.

地震荷载(earthquake loading)　结构力学的基本概念之一. 指由地震地面运动加速度 ($\ddot{x}_0(t)$) 使结构物产生的惯性力、惯性力矩的效应. 它主要取决于地震级度、场地条件、结构物重量及其动力特性. 由于地面运动的多向性和随机性，作用于结构物上的荷载为随机荷载，一般可按多自由度体系或无限自由度体系计算结构的动力效应. 在工程中多采用基底剪力法和按振型组合的方法计算作用到各质点的地震力. 基底剪力法给出的地震力(Q)是由单自由度体系的弹性地震反应推导出来的. 一般地，阻尼系数 ζ 很小时，

$$Q = k \cdot \beta \cdot W.$$

其中 $k = (\ddot{x}_0|_{\max})/g$ 为地面最大加速度与重力加速度的比值，称为水平地震系数；

$$\beta = \frac{2\pi}{T} \cdot \int_0^t \frac{\ddot{x}_0(\tau)}{\ddot{x}_0|_{\max}} e^{-\zeta\omega(t-\tau)} \sin\omega(t-\tau)\,d\tau$$

为单自由度体系在地震力作用下最大加速度与地面运动最大加速度的比值，称为动力放大系数. 式中 T 为结构自振周期；$x_0(\tau)$ 为在时间为 τ 时的地面加速度；

$$\omega = \sqrt{\frac{k}{M}},$$

k 为结构刚度，M 为质量；W 为建筑物总重量. 地震力沿高度分布与振型曲线的形式质量高度分布有关. 按振型组合方法计算地震力时，对于第 j 振型中第 i 个质量的等效动荷载为

$$p_{ij} = m_i a_{ij} \gamma_j \omega_j \int_0^t \ddot{x}_0(\tau) e^{-\zeta\omega_j(t-\tau)} \sin\omega_j(t-\tau)\,d\tau,$$

反应的最大值

$$p_{ij}|_{\max} = W_i a_{ij} \gamma_j \alpha_j,$$

式中 m_i 与 W_i 分别为第 i 个质点的质量和重量；a_{ij} 为第 j 振型在第 i 个质点的位移，

$$\gamma_j = \sum_{i=1}^{n} m_i a_{ij} \Big/ \sum_{i=1}^{n} m_i a_{ij}^2$$

为第 j 振型的振型参与系数；$\alpha_j = k \cdot \beta_j$ 为第 j 振型的地震影响系数，β_j 为第 j 振型动力放大系数. 地面垂直方向的振动加速度峰值均为水平方向加速度峰值的 1/3～2/3 左右. 而在震中区附近，垂直方向振动比较激烈. 垂直方向振动的频带也有一定的宽度，其振动频率一般均高于水平方向的振动频率.

地震反应谱(earthquake response spectrum)　结构动力学的基本概念之一. 是具有不同的自振频率 ω 和临界阻尼比 ζ 的线弹性单质点体系，在给定地震水平加速度 $\ddot{x}_0(t)$ 作用下的最大反应函数. 在同样地面加速度 $\ddot{x}_0(t)$ 作用下，自振频率 ω(或周期 T) 不同的单质点反应是不同的时间函数，对每一地震加速度，各质点对地面的相对位移的最大值 $S_D=$

x_{max}与周期 T 之间的关系曲线为相对位移反应谱.它是周期 $T=2\pi/\omega$ 和 ζ 的函数.取多次地震记录的 S_D-T 曲线的包线或平均线，加以修正和调整后，即得规范中规定的位移反应谱曲线.在结构抗震设计中，广泛采用的是地震荷载的概念，但某些结构还可以采用能量的概念进行设计，这样还需考虑速度和加速度反应谱.速度反应谱

$$S_r=\dot{x}_{max}=\omega S_D$$

表示速度的最大值，式中 $S_D=\sqrt{A^2+B^2}$ 为振幅，A 与 B 分别用下式表示：

$$A(t,\omega)=\int_0^t e^{\zeta\omega(t-\tau)}\ddot{x}_0(\tau)\cos\omega\tau d\tau,$$

$$B(t,\omega)=\int_0^t e^{-\zeta\omega(t-\tau)}\ddot{x}_0(\tau)\sin\omega\tau d\tau.$$

加速度反应谱 $S_A=(\ddot{x}+\ddot{x}_0)_{max}=\omega S_r$ 表示质点相对地面运动加速度最大值和地面运动加速度最大值之和.地震时，作用在结构上的最大剪力

$$Q=kx_{max}=k\frac{1}{\omega}S_r=\sqrt{kM}S_r,$$

其中 k 为结构刚度.作用在单质点结构上的最大惯性力，即地震力

$$p_i=M(\ddot{x}+\ddot{x}_0)_{max}=MS_A=\sqrt{kM}S_r,$$

M 为质量.在规范中，采用地震系数 $\alpha=S_A/g$，就是以重力加速度 g 为单位的加速度反应谱.

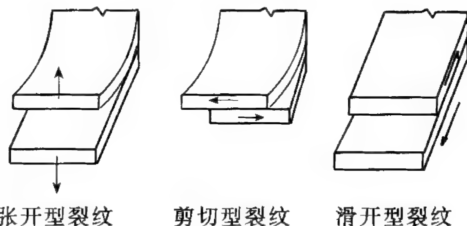
结构抗震分析(aseismatic analysis of structure)　抗震设计的主要环节.是对结构在预期地震作用下的动力反应的分析和计算.地震时的地面运动是一种非平衡的随机振动，其振动轨迹相当复杂，对不同的地震、不同的场地条件，甚至在同样场地条件下震中在不同的地点，地面振动都不一致，又受结构动力特性的影响，使计算极为困难.为了工程中的应用，必须作极大的简化.目前规范中规定的抗震结构分析，大致有静力法、基底剪力法、振型组合法和直接动力计算法等.在抗震设计中有多种模糊因素，所以近年来又提出抗震结构模糊优化设计的理论和方法.

断 裂 力 学

线弹性断裂力学(linear elastic fracture mechanics)　断裂力学的一个分支.是用线性弹性理论研究物体裂纹扩展规律和断裂准则的科学.线弹性断裂力学可用来解决脆性材料的平面应变断裂问题，适用于大型物体和脆性材料的断裂分析.1921 年，格里菲思(Griffiths,A.A.)通过分析材料的低应力脆断，提出裂纹失稳扩展准则——格里菲思准则.1957 年，欧文(Irwin,G.R.)通过分析裂纹尖端附近的应力场，提出应力强度因子的概念，建立了以应力强度因子 K_I(下标 I 表示张开型断裂)为参数的裂纹扩展准则，后来又有应变能释放率 G 作为线弹性断裂参数来判断断裂问题.裂纹尖端附近的应力超出弹性比例极限进入塑性阶段而成一塑性区，若该区很小(如远小于裂纹长度)，则可近似采用线弹性断裂力学分析；但对于韧性材料，不符合平面应变要求(如薄板)的构件及外部环境引起非线性的断裂问题，不宜采用线弹性断裂力学分析，而应采用弹塑性断裂力学来分析.线弹性断裂力学已发展得比较成熟，但还存在一些问题，如表面裂纹分析、复合型断裂准则、裂纹动力扩展等，有待进一步研究.

弹塑性断裂力学(elastic-plastic fracture mechanics)　断裂力学的一个分支.它应用弹塑性力学研究物体裂纹扩展规律和断裂准则，适用于裂纹尖端附近和裂纹体内部已有较大范围的塑性区的情况.裂纹尖端附近应力场十分复杂，求它的塑性区断裂问题的解析解非常困难，因此多采用近似方法，如 J 积分法、COD 法、R 曲线法或用实验方法进行分析.对薄板平面应力断裂问题的研究，也采用弹塑性断裂力学理论.该理论在焊接结构的缺陷评定、压力容器、飞行器、核电工程的安全评定研究等方面起重要作用.该理论仍不成熟.

裂纹(crack)　断裂力学的基本概念之一.指固体中破坏其连续性并且有尖锐几何形状的空隙.材料中的夹杂物、缩孔、表面划伤等都可算作裂纹.按裂纹在结构中的位置，可分为穿透裂纹、表面裂纹及埋藏裂纹.依裂纹扩展的方式，可分为张开型(I型)、剪切型(Ⅱ型)、滑开型(Ⅲ型)等三种基本型式.物体中任一裂纹扩展形式，都超不出这三种基本类型或它们的某种组合.



张开型裂纹　　剪切型裂纹　　滑开型裂纹

应力强度因子(stress intensity factor)　表征材料断裂的重要参量.是表征外力作用下弹性物体裂纹尖端附近应力场强度的一个参量.1957 年，欧文(Irwin,G.R.)建立了以应力强度因子为参量的裂纹扩展准则——应力强度因子准则，从而成功地解释了低应力脆断事故.张开型(I型)裂纹应力强度因子 K_I 是线弹性断裂力学中一个重要断裂参量.设外载和结构均以裂纹 $2a$ 为对称面，在裂纹尖端取极坐标为参考坐标，根据弹性力学给出的裂纹尖端附近的应力场，应力分量与应力强度因子的关系可近似写为

$$\sigma_x = \frac{K_{\mathrm{I}}}{\sqrt{2\pi r}} \cdot \cos\frac{\theta}{2}\left(1 - \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right),$$

$$\sigma_y = \frac{K_{\mathrm{I}}}{\sqrt{2\pi r}} \cdot \cos\frac{\theta}{2}\left(1 + \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right),$$

$$\tau_{xy} = \frac{K_{\mathrm{I}}}{\sqrt{2\pi r}} \cdot \cos\frac{\theta}{2}\sin\frac{\theta}{2}\cos\frac{3\theta}{2},$$

其中 σ_x 和 σ_y 分别为平行和垂直于裂纹面的正应力，τ_{xy} 为剪应力分量；r,θ 为极坐标，其原点为裂纹尖端. Ⅱ型和 Ⅲ型裂纹也有类似的应力强度因子 K_{II} 和 K_{III}.

疲劳裂纹扩展速率(fatigue crack growth rate) 断裂力学概念. 指交变应力每循环一次裂纹长度的增加量. 通常用 $\mathrm{d}a/\mathrm{d}N$ 表示，式中 a 为裂纹长度，N 为应力循环次数. 循环特性一定时，可用帕里斯公式估算：

$$\frac{\mathrm{d}a}{\mathrm{d}N} = c(\Delta k)^m,$$

式中 Δk 为应力强度因子变程，c 和 m 为材料常数. $\mathrm{d}a/\mathrm{d}N$ 对于估算裂纹体疲劳寿命有重要作用.

临界应力强度因子变程(critical value of alternating stress intensity factor) 亦称裂纹扩展门槛值. 表征材料制止裂纹扩展的能力. 是指裂纹基本上不发生疲劳扩展的应力强度因子变程的上限值. 通常用 Δk_{th} 表示. 裂纹体承受交应力时，只要其应力强度因子变程 $\Delta k \leqslant \Delta k_{\mathrm{th}}$，则裂纹不会扩展，即

$$\frac{\mathrm{d}a}{\mathrm{d}N} = 0.$$

循环特性一定时，门槛值 Δk_{th} 是一材料常数.

裂纹扩展门槛值(threshold value of crack growth) 即“临界应力强度因子变程”.

应力腐蚀临界应力强度因子(critical value of stress intensity factor by stress-corrosion) 表征材料制止裂纹扩展的能力. 是裂纹在腐蚀介质中不发生扩展的应力强度因子的上限值. 对于 I 型裂纹，通常用 K_{ISCC} 表示. 对于一定材料，K_{ISCC} 一般低于 K_{IC}. 材料的 K_{ISCC} 主要取决于腐蚀介质，且与化学成分及金相组织有关.

应变能释放率(strain energy release rate) 亦称裂纹扩展力. 线弹性断裂力学中一个重要断裂参量. 是含有裂纹物体的应变能随裂纹扩展的变化率. 1921 年，格里菲思(Griffiths，A. A.)建立了以应变能释放率为参量的裂纹失稳扩展准则——格里菲思准则，从而成功地解释了玻璃的低应力脆断现象. 应变能释放率 G(恒大于零)是表示裂纹扩展单位面积系统所能提供的能量. 其表达式(对于 I 型裂纹)为

$$G_{\mathrm{I}} = \frac{\partial U}{\partial A},$$

式中 A 为切开裂纹面的面积，U 为构件的应变能. G_{I} 与 K_{I}(张开型裂纹应力强度因子)之间存在一定关系：

$$G_{\mathrm{I}} = \frac{K_{\mathrm{I}}^2}{E}\text{(平面应力问题)；}$$

$$G_{\mathrm{I}} = \frac{K_{\mathrm{I}}^2(1-\nu)}{E}\text{(平面应变问题)，}$$

E 为材料弹性模量，ν 为泊松比.

裂纹扩展力(force of crack growth) 即“应变能释放率”.

断裂参数(fracture parameter) 描述含有裂纹物体断裂特性的一些物理量. 断裂参数具有以下特性：能通过实验测定；能通过理论算出；对应不同类型的断裂，当相应的参数达到临界值时，含裂纹物体就发生断裂，参数的临界值仅取决于材料性质，与外载和物体的几何形状无关. 目前常用的断裂参数有应力强度因子 K_{I}(及 K_{II}，K_{III})、应变能释放率 G，J 积分、裂纹张开位移(COD) 等. K_{I}(及 K_{II}，K_{III}) 和 G 是用于线弹性断裂力学的断裂参数，具有严密的理论基础，成功地揭示了裂纹体的断裂规律，国际上已有 K_{I}(及 K_{II}，K_{III}) 计算手册. J 积分、COD 是用于弹塑性断裂力学的断裂参数. J 积分具有一定的理论基础，COD 已广泛用于压力容器的缺陷评定标准. 复合型判据已于 20 世纪 80 年代中期得到广泛重视，是亟待解决的问题. 物体断裂时，断裂参数的临界条件称为断裂准则，不同的断裂参数对应于不同的断裂准则.

断裂韧度(fracture toughness) 材料抵抗裂纹扩展的能力. 定量描述材料断裂韧性好坏的参量称为断裂韧度. 确定材料韧性参量传统的实验方法有：却贝 V 型缺口、落锤、动态撕裂、赛板、三点弯曲实验等，这些在一定程度上模拟结构实际工作状态，定量地描述材料韧性的好坏. 不同的理论，提供不同的断裂韧性指标和测试方法. 断裂力学中常用的断裂韧性指标有：临界应力强度因子 K_{IC}、临界 J 积分值 J_{IC}、临界裂纹张开位移 δ_{C} 等. K_{IC} 适用于线弹性材料，J_{IC}，δ_{C} 适用于弹塑性材料，它们的数值反映材料抵抗断裂能力的大小. 影响材料断裂韧性的因素除化学成分及热处理工艺外，还有环境温度、加载速率等. 进行结构设计时，必须同时考虑强度指标和断裂韧性指标.

应力强度因子准则(stress intensity factor criterion) 以应力强度因子为参量的裂纹失稳扩展准则. 其内容为：张开型裂纹扩展的临界条件为 $K_{\mathrm{I}} = K_{\mathrm{IC}}$. 其中 K_{I} 为张开型应力强度因子，可由弹性力学方法求得；K_{IC} 为材料的临界应力强度因子或平面应变断裂韧度，可由实验测定. 平面应变断裂韧度是反映物体断裂特性以及工程安全设计的重要参量，它的测定是断裂力学的一项重要任务.

复合型断裂准则(mixed mode fracture criterion)　关于断裂的判定准则.指判断裂纹体在多种裂纹扩展类型同时存在的情况下是否断裂的准则.由于物体中应力场同原裂纹方向不对称以及材料的各向异性,从而产生复合型变形状态,出现不同形式的裂纹组合,即为不同形式的复合裂纹.如Ⅰ型和Ⅱ型同时存在,应力强度因子为$K_{(\mathrm{I}+\mathrm{II})}$,或有其他形式的组合,则有相应于这些类型裂纹的应力强度因子.不同的复合型断裂,就是不同的应力强度因子存在下的断裂现象,Ⅰ,Ⅱ,Ⅲ型复合型断裂准则,就是寻求$K_{(\mathrm{I}+\mathrm{II}+\mathrm{III})}$的一个函数关系式

$$f(K_{(\mathrm{I}+\mathrm{II}+\mathrm{III})}) = 0,$$

当此关系式成立时,裂纹就扩展.在复合型断裂中,裂纹一般不沿裂纹原来方向扩展,其裂纹扩展方向同原裂纹方向有一夹角(断裂角),复合型断裂准则应确定出现断裂角以及在扩展方向上断裂参量$K_{(\mathrm{I}+\mathrm{II})}$(如Ⅰ,Ⅱ型参量)的临界值.目前常用的复合型断裂准则有能量释放率准则、最大法应力准则和应变能密度因子准则.它们确定的断裂角与实验比较符合,但确定的断裂参量临界值与脆性材料的实验结果有较大差距,特别是当$K_{\mathrm{I}}/K_{\mathrm{II}}$较小时,理论值与实验值相差2～3倍.

平面应变断裂韧度(plane strain fracture toughness)　材料断裂韧性指标.平面应变断裂的条件为裂纹尖端塑性区尺寸远小于裂纹长度和其他几何尺寸.它出现在试样厚度较大的情况中.为保证平面应变状态的实现,试样尺寸应满足:

$$B, a, (W-a) \geqslant 2.5(K_{\mathrm{IC}}/\sigma_s)^2,$$

其中W,B分别为试样宽度和厚度;a为裂纹长度;σ_s为屈服应力;K_{IC}为张开型临界应力强度因子或平面应变断裂韧度,可由三点弯曲实验、紧凑拉伸实验等方法测定.

裂纹张开位移(crack opening displacement)　简称COD.弹塑性断裂力学中的一个断裂参量.是含裂纹的物体,裂纹上下表面受拉伸荷载作用下在原裂纹尖端出现张开型的相对位移.它是用来表征含裂纹的平板在裂纹尖端的物理状态.

裂纹张开位移法(crack opening displacement method)　判定裂纹扩展的方法.指弹塑性断裂力学中以裂纹张开位移作为断裂参量判别裂纹开始扩展的一个近似的工程方法.该法由韦尔斯(Wells, A. A.)于1963年提出.当裂纹尖端张开位移δ达到临界值δ_{C}时,不管含裂纹体的形状、几何尺寸、受力大小和方式如何,该裂纹即开始扩展.δ_{C}是反映材料断裂性能的常数,可由实验得到.对于韧性材料,平面应力断裂问题,特别是裂纹体内出现大范围屈服,甚至完全屈服等线弹性断裂力学理论不能适用的情况下,可采用此法.该法比较简单、直观,已反映在压力容器缺陷的评定标准中.在线弹性范围内,裂纹张开位移与Ⅰ型裂纹应力强度因子K_{I}之间的关系为

$$\delta = K_{\mathrm{I}}^2 \cdot \frac{1-\nu^2}{E\lambda\sigma_s},$$

与J积分之间的关系为$J=k\sigma_s\delta$,式中ν为材料的泊松比,E为材料弹性模量,σ_s为屈服应力,λ,k均为常数.

J积分(J integration)　弹塑性断裂力学中的一个重要断裂参量.是在含有裂纹的物体内围绕裂纹尖端的一个与路径无关的回路积分.1968年,赖斯(Rie, J. R.)定义的二维含裂纹J积分为

$$J = \int_\Gamma \left(W\mathrm{d}y - \boldsymbol{T} \cdot \frac{\partial \boldsymbol{u}}{\partial x}\mathrm{d}s\right),$$

式中x,y分别表示以裂纹尖端为坐标原点、平行和垂直于裂纹面的坐标轴,W为非线性弹性体的应变能密度,Γ为围绕裂纹尖端逆时针方向的任意积分回路,$\mathrm{d}s$为Γ上的微段弧长,$\boldsymbol{T}$为作用在$\mathrm{d}s$上的张力矢量,$\boldsymbol{u}$为相应的位移矢量.J积分用来描述裂纹尖端附近在非线性弹性情况下的应力场.当J的值达到极限值J_{IC}时,裂纹开始扩展.J_{IC}是表征材料断裂韧性的临界J积分值,它可由实验确定.近年来,J积分已被推广应用到三维非线性弹性体的有限变形问题中,还被用来进行蠕变和疲劳裂纹扩展的分析.目前发展出按照弹性和塑性两种极限状态下计算J积分的近似方法,并编制出典型试件和含裂纹简单构件的J积分计算图表.J积分作为弹塑性断裂参量,有它一定的近似性,对于有卸载情况的断裂问题不宜采用.对于有限变形和考虑温度时,J积分表达式要作相应修改.

撰　稿　丁树人　任玉安　杨亚政　李　星　沈钧涛　陈守中　武际可　金孝维

审　阅　刘延柱　杨桂通　李万琼　张善元　姚振汉　黄文彬　蒋持平　戴世强

天　文　学

天文学(astronomy)　现代自然科学基础学科之一.它研究宇宙中的“大尺度”目标和现象.按照尺度的规模,可以分为四个层次:

1.地球、太阳系行星和行星系统(包括太阳系小天体和行星际物质).

2.太阳、各类恒星和恒星系统(包括恒星际物质).

3.银河系、各类星系和其他河外天体以及星系群、星系团等系统(包括星系际物质).

4.“宇宙”整体.

下面这些典型数据,可以代表目前对各个不同层次的尺度所作的估计:地球直径为 1.3×10^{-11}光年;太阳直径为 1.6×10^{-8}光年;太阳系范围约 10^{-4} 光年;最近恒星距离约 4 光年;银河系范围约 10^5 光年;邻近星系典型距离约 10^6 光年;“富”星系团大小约 10^7 光年;可测宇宙范围约 2×10^{10}光年(1 光年约为 9.46×10^{12}千米).

天文学中的“宇宙”是一个有机的整体.星系层次的天体是大尺度宇宙中的“细胞”,而恒星层次的天体则既是星系中的“细胞”,又是行星层次的“母体”.所有的天体都在运动中保持它在群体中的相对平衡,又在演化中不断改变它的内在结构和在群体中的作用.

地球是行星层次中的一个天体,属太阳系中九大行星的一员,它在浩瀚无边的宇宙中犹如沧海里一粒微尘.地球本身是迄今所知的惟一繁衍着生命的星球.太阳,作为太阳系的主体,是离地球最近、对人类影响最大的一颗恒星,尽管它在恒星世界中毫不突出,但却是人们惟一得以详加研究的这一尺度的目标.太阳带着太阳系,以及居住在地球上的人们,处在一个“旋涡星系”——银河系里,它离银河系中心约 3 万光年,以 250km/s 的速度绕银河系中心飞奔.银河系的形状像一个中央隆起的圆盘,盘的直径约 10 万光年,其中稀疏地散布着约 1000 亿个恒星,以及极为稀薄、但是成团成块地弥漫在整个空间中的星际物质.有证据表明,所有的恒星都是由星际物质吸积而成,而各类恒星的一生中都在不停地向星际空间抛射物质.抛射的方式大部分是持续而比较缓慢的,但有的表现为不同程度的“激变”;相当多的恒星在“正常”生命结束的刹那,会发生剧烈的爆炸,把自身物质的大部分猛烈地抛到周围空间.

银河系好比是广阔宇宙中的一个“岛”.人们住在这个“岛”上,可以就近看到它的内部结构和活动,但却难以了解它的全貌.银河系以外的星系,近一些的离地球约 100 万光年,远的达 100 亿光年以上.用现代“巡天”望远镜可以观测到的星系不下 10 亿个.它们在空间中常结成团,典型的“富”星系团含有数以千计的星系,占据空间近千万光年.

在星系层次中,人们观测到宇宙间极其强烈的能量爆发和规模巨大的物质抛射,其中离地球很远的类星体和活动星系,发出的辐射可超过一般星系辐射的千百倍,有的抛出的物质可以达几百万个太阳质量.

前面是对人们今天“看”到的天文世界的表观描述.这些表观现象的认识来自天文观测和观测资料的处理和分析.而从表观的认识提高到本质的认识,则要借助于物理学知识.天文观测、处理分析和物理解释三个步骤,以及三者之间的交叉环节,可以认为是天文学发展的内因.

天文实测的发展取材于同时代的技术发展;天文现象的物理解释以同时代的物理学为依据;而同时代数学的发展则时刻渗透在物理解释和数据的处理、分析中.在这种意义上,同时代的技术、物理学和数学的发展是天文学发展的直接外因.

天文观测的对象是天体投来的“光”(有一些例外,如中微子、宇宙线粒子、陨星等,但份量很少).这里用带引号的“光”泛指电磁波谱中任何辐射,因为尽管迄今绝大部分的天文信息是由可见光带来的,但射电、X 射线、红外等波段的信息却往往联系到天体的不寻常表现.

宇宙之大,天体之多,每深入一步研究,都要求测量到更暗弱以及更遥远(因而也成为更暗弱)的目标.这种要求没有止境,因而对天文望远镜的聚“光”、感“光”能力的要求也没有止境.同样的理由,对天文望远镜测量天体位置的精度、分辨天体细节的精度以及辨认天体的位移和“光变”的精度,也同样有着无止境的要求.天文学的无止境要求是天文仪器(和技术)发展的恒在动力,而随着天文仪器的发展接踵而来的天文发现,又为天文学的发展提供了持续的推动力.

这种良性循环以天文仪器(和技术)的发展为“主线索”.在现代天文学萌芽之前,构成这种循环的中间环节是“惟象模型”的建立(也就是前面提到的对天文资料的处理、分析,并由此获得的经验规律),而天文学进入到现代的重要转折是由“惟象模型”提高到“物理模型”(也就是前面提到的对天文观测现

象或经验规律的物理解释).第谷—开普勒—牛顿时期的天文学进程标志着这一转折,产生了天文学历史上一次巨大飞跃.

这一时期机械技术的发展足以使第谷(Tycho Brahe)把天体定位的精度提高到几十[角]秒,他以前所未有的水平进行了行星和恒星的定位观测.接着是开普勒(Kepler,J.),他处理、分析了第谷的观测数据,总结出了开普勒三定律.然后是牛顿(Newton,I.),他根据开普勒定律和地面的力学实验结果,发现了万有引力,建立了力学定律,同时创建了用力学来研究天体运动的学科——天体力学.天体力学的诞生使天文学从以往单纯探讨天体运行的经验关系,进而研究天体间的相互作用.这种研究对其后的天文学、力学和数学的发展起着巨大的影响,对行星层次的认识,则是一次从现象到本质的提高.

天文学进入20世纪,同时期的光学、机械以及照相、计时、分光等技术的发展,使天文学家得以拥有口径为1m级望远镜,积累了大批恒星的光度资料和一批恒星光谱资料.这时的恒星定位精度可达0.01″级,利用三角视差方法测定了数以千计恒星的距离.这些同时代技术的引进和大量观测资料的积累,可以比拟于前面所说的第谷的工作.这些资料的处理、分析使赫茨普龙(Hertzsprung,E.)和罗素(Russell,H.)能分别通过各自的方法,得到了恒星表面温度和光度之间的统计关系,即后来以他们的姓氏命名的赫-罗图.赫-罗图可以认为是"开普勒型"的工作,它的发现使理论天体物理学家们能够结合经典物理以及当时发展的原子物理和核物理,建造出恒星内部结构模型以及以恒星中心核聚变为线索的恒星演化理论,这个理论从最简单的星云学说出发,一举解释了表面截然不同的两大自然现象,即各种类型恒星的起源和演化以及周期表中所有元素的合成.这个理论还预言了中子星和黑洞的存在.用前面的比拟,这一部分工作属于"牛顿型".这一轮"第谷型—开普勒型—牛顿型"工作的接力,使人们对恒星层次的认识得到了从现象到本质的提高.

同是在20世纪上叶,星系和宇宙层次也出现了一次从现象到本质的认识飞跃.黑尔(Hale,G.E.)发挥当时光学和机械技术的能力,成功地研制成直径2.5m望远镜.以此为代表的大、中型天文设备首先确定了许多"星云"是河外星系,继而测量了大批河外星系的光度、形态和光谱.1929年,哈勃(Hubble,E.P.)处理、分析这些资料,提出了星系距离与谱线红移成正比的哈勃定律.这是一条经验(惟象)定律,但是当它很自然地结合引力论(首先是广义相对论),构成以膨胀宇宙为出发点的现代宇宙学,特别是应用亚原子物理(核物理、基本粒子物理等)发展起来的热大爆炸宇宙学模型,哈勃定律成为其中的主要实测依据,从而融入了这个物理模型,本身也成为物理规律.热大爆炸宇宙学把微观世界和天文世界的起源与演化统一到同一个物理模型里,可以把今天物质世界中一切基本结构,例如,人体中的分子结构的元素成分——C,O等追溯到宇宙"初生"一刹那的基本过程.它是今天被较多人采用的"工作模型",具有解释以及预测更多的观测事实的功能.发挥这种功能可以促使模型的自身改善和应用范围的开拓,同时也是承受实测以及理论检验和挑战的"头号靶子".

20世纪60年代以来,恒星演化模型和热大爆炸宇宙学模型一直不断地追求新的开拓,接受着新的检验,并遭遇到新的挑战.在这期间,成为当代天文学发展"主线索"的天文仪器(和技术)取得了革命性的成就.传统的光学望远镜,在孔径5m的大型反射镜独步一时之后,陆续在全球各地配置了近10台孔径4m级望远镜;感光装置,从照相、光电、电真空系统发展到量子效率近于理想的电荷耦合探测器.到20世纪80年代末,用灵敏感光器件配备的大型望远镜,辨察微弱天体的能力已超过肉眼一百亿倍.然而,这一期间天文仪器划时代的进展却应首推射电天文和空间天文手段的发展与成熟.这两者与传统光学手段鼎足而三,共同开创了全波段天文观测的新纪元.20世纪60年代,初具规模的射电天文手段带来的四大天文发现曾经震动科学界.它们中间,脉冲星和微波背景辐射的发现分别证实了恒星演化理论和热大爆炸宇宙学的预言,对这些物理模型提供了决定性的支持;类星体和星际复杂分子的发现则分别开拓了星系层次和恒星层次的新领域.接着从20世纪70年代跨进到20世纪90年代,以IUE紫外卫星、"爱因斯坦"X射线卫星、IRAS红外卫星为代表的空间手段,以及一批大型射电天文设备(包括有效口径100m级的厘米波望远镜和干涉阵,口径十几米到四十几米的毫米波望远镜,以及分辨能力达万分之几(角)秒的甚长基线干涉仪等)接连投入工作,各个层次新领域的开拓进展迅速,新的重要发现源源涌现.观测资料和结果的这种高速积累,使人们感到天文学在新的一轮循环中的理论突破正在趋近.20世纪90年代中的地面光学、射电天文以及空间天文设备的聚"光"能力较20世纪80年代全面提高一至几个量级;太阳和行星际空间以及日地关系的研究也由于大型太阳望远镜的投入和一系列专用卫星的发射而得到新的进展;20世纪90年代中,宇宙飞船远访行星的计划进一步进行;几个大型中微子探测器在各地建成.这些使人们预计五至十年后的天文学将需要有新的描述,特别是摆在眼前的一系列重要发掘和严峻挑战,如恒星和行星形成的红外、亚毫米波探索,太阳中微子"失踪"之谜,地外

智能寻觅，黑洞寻觅，星系、星系团动力学结构中占质量 90%的"暗物质"的寻觅，类星体、活动星系巨大产能机制的探索，微波背景辐射极端均匀与星系群体大尺度分布不均匀的矛盾，星系演化等，人们正期待着了解这些探索的结果和"必然出现的偶然发现"将如何丰富或者如何改建天文世界现有的各类模型.

天体测量学

天体测量学(astrometry) 天文学的一个分支学科. 主要研究精确测定天体的位置和运动、建立基本参考坐标系、精确测定地球表面点位的坐标和地球自转速率的理论和方法，以及确定基本天文常数系统. 它是天文学中发展最早、应用最广的一个分支. 具体可分为：

1. 球面天文学，研究天体的天文坐标的表达方式和换算. 坐标需加的各种改正，是天体测量学的理论基础.

2. 方位天文学，研究天体位置和运动的测定、参考坐标系的建立.

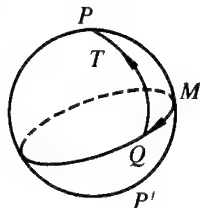
3. 实用天文学，研究时间和纬度的测定、天文经纬度和方位角的测定以及航海、航空的天文导航等，与大地测量学、制图学关系密切.

4. 天文地球动力学，研究地球自转、地极移动、地壳运动的测定和机制.

20 世纪 60 年代以后，由于新技术，如激光、射电天文技术和空间技术等的应用，使天体测量学有了新的发展. 观测方式已由测量角度扩展到测量距离；观测所在地已由固定天文台发展为流动站、全球性组网观测以及空间观测；观测精度大为提高，测角精度达 0.001″，测距精度达厘米级；并逐渐形成射电天体测量学、空间天体测量学一些新的分支.

球面天文学(spherical astronomy) 天文学的一个分支学科. 它运用球面三角学和矩阵运算等数学方法研究天体在天球上的投影位置及其变化，是天体测量学、天体力学、恒星天文学等分支学科的理论基础之一. 它的研究内容主要有：天球坐标系的建立，天体因地球自转和公转造成的天体的视运动，因地球运动使观测者位于空间不同位置而造成的视差对天体视位置的影响，大气折射理论及其对天体视位置的影响，地球运动和光速有限造成的光行差理论及其对天体视位置的影响，因天体引力摄动造成的地球公转轨道面(黄道面)和地球赤道面的变化引起的天体视位置的变化(岁差和章动)，以地球自转和公转为基础的时间计量系统(世界时和历书时)的建立以及天体视位置的归算方法等.

天文坐标系(astronomical coordinate system) 亦称天球坐标系. 一种坐标系. 用来确定天体在天球上的位置的球面坐标系称为天文坐标系. 选取天球上某一大圆作基圈，基圈的极点作该球面坐标系的基本点，过两极的大圆均与基圈垂直，选取其中一个大圆作主圈，以基圈与主圈的交点作主点，其余大圆称副圈. 表示天体的两个球面坐标分别是：



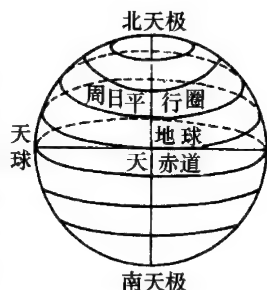
1. 在通过天体 T 的副圈上，从基圈与副圈的交点 Q 到天体间的弧段 QT(纬度)或极点与天体间的弧段 PT(矩).

2. 主点 M 沿基圈到过天体 T 的副圈与基圈的交点 Q 间的弧段 MQ(经度).

根据不同需要，有地平坐标系、赤道坐标系、黄道坐标系、银道坐标系等.

天球坐标系(celestial coordinate system) 即"天文坐标系".

天球(celestial sphere) 天体测量学中的一个概念. 以观测者为球心，以任意长(无穷大)为半径的假想球面称为天球. 由于天体距离非常遥远，看起来它们好像分布在一个巨大的圆球面上. 观测者实际上所看到的是天体在天球上的投影点. 天文学中常用这些投影点之间的大圆弧来表示天体的位置和视运动的数量. 由于地球自转，看起来天球好像围绕假想的天轴每天自东向西旋转一周，这种运动称为天球周日视运动. 天轴实际是地球自转轴延长与天球相交的假想轴线，地球自转轴延长与天球的两个交点称天极. 地球北极向上的交点称北天极，另一点称南天极. 地球赤道无限扩大与天球相交的大圆称天赤道.



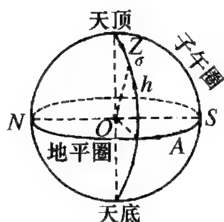
天轴(celestial axis) 见"天球".

北天极(north celestial pole) 见"天球".

南天极(south celestial pole) 见"天球".

地平坐标系(horizontal coordinate system) 一种天文坐标系. 以地平圈为基圈、天顶为基本点的天文坐标系称为地平坐标系(参见"天文坐标系"). 如图，过观测者 O 的铅垂线与天球的交点，向上的交点称天顶，向下的交点称天底. 以子午圈为主圈，以南点 S 或北点 N(子午圈与地平圈相交于两点，距北天极近的称北点，另一点称南点)为主点. 表示天体位置的两个地平坐标分别为地平高度 h(又称地平纬度)或天顶距 Z、方位角 A(又称地平经度). 地平高度是自地平圈沿通过天体 σ 的地平经圈(过

天顶、天底与地平圈垂直的大圆)到天体的弧段,自地平圈向天顶和天底量度,从 0°到 90°,在地平圈以上为正,在地平圈以下为负. 地平高度的余角称天顶距,自天顶量起,从 0°到 180°. 方位角 A 是南点到通过天体的地平经圈与地平圈交点间的弧段,自南点沿地平圈按顺时针方向(从天顶看去)量度,从 0°到 360°. 在大地测量学中自北点按顺时针方向量度. 天体的地平坐标值随天体周日视运动而不断变化. 此外,观测者地理纬度不同,天体的地平坐标值也不相同.



天顶(zenith) 见"地平坐标系".

天底(nadir) 见"地平坐标系".

南点(south point) 见"地平坐标系".

北点(north point) 见"地平坐标系".

地平经圈(azimuth circle) 见"地平坐标系".

子午圈(meridian) 天球上的一个大圆. 指通过天极和天顶的天球大圆,也是通过观测点和地球南北极的地理经圈扩展后与天球相交的大圆,它是位于南北方向的地平经圈,又称南北圈. 中国古代用十二地支表示方向,子表示北,午表示南,故名. 子午圈是天体地平坐标的基本圈之一,子午圈的方位角为 0°和 180°. 有时把北天极以南的半个大圆称为午圈,把北天极以北的半个大圆称为子圈. 天体周日视运动时两次通过子午圈的状态或时刻,称为中天. 通过午圈为上中天,通过子圈为下中天.

地平圈(horizon) 天球地平坐标系的基圈. 通过天球中心而垂直于天顶和天底连线的平面与天球相交的大圆称为地平圈. 天文学上称为真地平圈或数学地平. 天与地的交线则称为视地平圈. 在陆地上,视地平圈有时在真地平圈上面,有时在真地平圈下面. 海面上,视地平圈总是比真地平圈低,是一和真地平圈平行的小圆.

卯酉圈(prime vertical) 天球上与子午圈正交的大圆. 指过天顶与东点和西点的天球大圆. 即位于东西方向的地平经圈,又称东西圈. 中国古代以十二地支表示方向,以卯表示东,以酉表示西,故名. 它和地平圈相交于东点和西点,卯酉圈的方位角为 90°和 270°.

方位角(azimuth) 亦称地平经度. 常称方位. 天体在地平坐标系中的一个坐标. 是子午圈与通过天体的地平经圈在地平圈上所夹的弧段,或子午圈与通过天体的地平经圈在天顶所夹的两面角. 在地平圈上自南点起顺时针方向(自天顶观看)到通过天体的地平经圈与地平圈的交点,或自天顶以南的子午圈顺时针方向到该天体的地平经圈的角度,从 0°到 360°. 大地测量、航海、航空则是自北点起顺时针方向到通过天体的地平经圈与地平圈的交点,或从天顶以北的子午圈顺时针方向到该天体的地平经圈的角度.

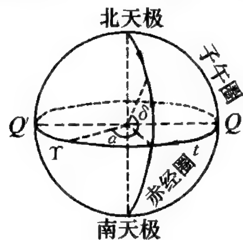
地平经度(azimuth) 即"方位角".

地平高度(altitude) 亦称地平纬度. 常称高度. 天体在地平坐标系中的一个坐标. 是从地平圈沿通过天体的地平经圈到天体的弧段,即观测者与天体的连线和地平面之间的夹角. 地平圈以上为正,以下为负,各自从 0°到 90°.

地平纬度(altitude) 即"地平高度".

天顶距(zenith distance) 天体地平坐标系中的一个坐标. 从天顶沿通过天体的地平经圈到天体间的弧段,自 0°到 180°. 天顶距 Z 与地平高度 h 互为余角,即 $Z=90°-h$.

赤道坐标系(equatorial coordinate system) 亦称第二赤道坐标系. 天文坐标系的一种(参见"天文坐标系"). 以天赤道为基圈,以北天极为基点,通过天极与天赤道垂直的大圆称赤经圈,以通过春分点的赤经圈为主圈,以春分点 ♈ 为主点. 表示天体位置的赤道坐标为赤纬 δ 和赤经 α. 赤纬是自天赤道沿通过天体的赤经圈到天体的弧段,自天赤道分别向北和向南计量,各从 0°到 90°,向北为正,向南为负. 赤纬的余角称为北极距,即自北天极沿过天体的赤经圈到天体的弧段,从 0°到 180°. 赤经是春分点和通过天体的赤经圈与天赤道的交点之间的弧段,自春分点沿天赤道按与天体周日视运动相反的方向量度,从 0～24h. 天体的赤道坐标不因观测地点改变而变化,也不随周日视运动而变化.



第二赤道坐标系(second equatorial coordinate system) 即"赤道坐标系".

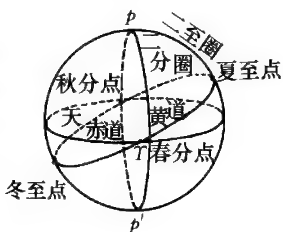
赤经圈(right ascension circle) 见"赤道坐标系".

天赤道(celestial equator) 天球大圆之一. 地球赤道面与天球相交的大圆称为天赤道. 天赤道将天球分成两个半球:南半天球和北半天球. 天赤道是天体赤道坐标系的基圈,天赤道的赤纬为 0°. 天赤道与黄道相交,交角为 23°26′,两交点为春分点和秋分点.

天极(celestial pole) 天球坐标系的基本点之一. 地球自转轴延长线与天球相交的两点称为天极. 自地球北极向上的那点称为北天极,另一点称为南天极. 天极是天球赤道坐标系的基点. 由于地球赤道隆起部分受到日、月的附加吸引,地球自转轴围绕黄

轴(通过地心垂直于黄道的轴线)旋转,使得天极位置在天球上绕黄极以 23°26′为半径,以约 25800a 的周期描绘出一个有微小波线起伏的圈.天极处的恒星在不同时期是不同的,目前北天极附近的恒星为小熊座 α 星(勾陈一),称它为北极星.8000a 后北天极则移到天鹅座 α 星附近.

春分点(vernal equinox) 天球坐标系的一个基本点.即天赤道与黄道的交点之一.每年 3 月 21 日前后,太阳沿黄道从天赤道以南经过此点运行到赤道以北.天文学上定此时为北半球春季开始.春分点是黄道坐标系、赤道坐标系计量黄经、赤经的起点.两千多年前,春分点在白羊座中,故以白羊座符号 ♈ 表示.由于日、月、行星的摄动(岁差和章动),春分点的位置不断变化,



因而有平春分点(只考虑岁差作等速运动的春分点)和真春分点(随岁差和章动而移动的春分点)之分.星表、星图所给出的天体位置均指对应于某一历元的天赤道和平春分点的位置.由于岁差,春分点现在移到双鱼座.天赤道与黄道的另一交点称为秋分点,两千多年前,秋分点位于天秤星座,故以天秤座符号 ♎ 表示,现在秋分点西移到室女座内.每年 9 月 23 日前后,太阳经该点从天赤道以北运行到天赤道以南,天文学上定此时为北半球秋季开始.秋分点与春分点合称二分点.经过春分点、秋分点和天极的赤经圈称二分圈.与春分点黄经相差 90°,270°的点分别称为夏至点和冬至点,每年 6 月 22 日前后、12 月 22 日前后,太阳分别经过夏至点和冬至点,各表示北半球夏季和冬季的开始.夏至点和冬至点合称二至点,连结二至点和天极的赤经圈称二至圈.

二分点(equinoxes) 见"春分点".

二分圈(equinoctial colure) 见"春分点".

二至点(solstices) 见"春分点".

二至圈(solstitial colure) 见"春分点".

赤经(right ascension) 天球赤道坐标系的一个坐标.指通过春分点的赤经圈与通过天体的赤经圈之间在天赤道上的弧段,以 α 表示.自春分点起,按与天球周日视运动相反的方向量度,从 0h 到 24h.

赤纬(declination) 天球赤道坐标系的一个坐标.指从天赤道沿通过天体的赤经圈到该天体的弧段,以 δ 表示.即观测者与天体的连线和天赤道之间的夹角.天赤道以北为正,天赤道以南为负,从 0°到 90°.

极距(polar distance) 赤道坐标系中赤纬的余角.指自北天极沿通过天体的赤经圈到天体间的角距离,以 p 表示,自 0°到 180°.它与天体赤纬 δ 的关系为 $p+\delta=90°$.对南半球的恒星,其极距

$$p=90°-(-\delta)=90°+\delta>90°.$$

时角赤道坐标系(dependent equatorial coordinate system) 亦称第一赤道坐标系或时角坐标系.天文坐标系的一种.坐标系的基圈为天赤道,基点为北天极.过天极与天赤道垂直的大圆称为时圈.以天球子午圈为主圈,以天赤道与午圈的交点 Q(称为上点)为主点.表示天体位置的两个坐标为赤纬 δ 和时角 t.赤纬 δ 定义为自天赤道沿通过天体的时圈到天体间的弧段,自天赤道分别向北和向南量度,从 0°到 90°,天赤道以北为正,天赤道以南为负.有时也用赤纬的余角,称为极距,即 $90°-\delta$,定义为自北天极沿过天体的时圈(即赤经圈)到天体的弧段,从 0°到 180°.时角 t 定义为过天体的时圈与天赤道的交点和上点间的弧段,自上点沿天赤道按天体周日视运动方向量度,从 0°到 360°,或从 0h 到 24h.当天体作周日视运动时,天体的赤纬不变,时角从 0h 均匀变化到 24h.

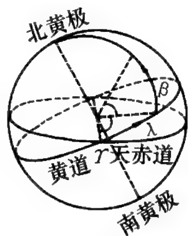
第一赤道坐标系(dependent equatorial coordinate system) 即"时角赤道坐标系".

时角坐标系(hour angle coordinate system) 即"时角赤道坐标系".

时圈(hour circle) 见"时角赤道坐标系".

时角(hour angle) 时角坐标系的一个坐标.指子午圈和通过天体的时圈两者在天赤道上所夹的弧段或在天极处的夹角,以 t 表示.量度方法是:自午圈与天赤道相交的交点(上点)起,沿天赤道到通过天体的时圈在天赤道上的垂足.按天体周日视运动方向量度,自 0h 到 24h,即自 0°到 360°.或者自午圈沿天体周日视运动方向到过天体的时圈,自 0h 到 12h,为正值;自午圈沿与天体周日视运动相反方向到过天体的时圈,自 0h 到 −12h,为负值.天体的时角 t 与天体赤经 α 之和在数值上等于春分点 ♈ 的时角 $t_♈$,即 $t_♈=\alpha+t$.

黄道坐标系(ecliptic coordinate system) 天文坐标系的一种.指以黄道为基圈,以北黄极(距黄道 90°的点称黄极,距北天极近的称北黄极,另一点称南黄极)为基点的天文坐标系.以通过春分点的黄经圈(通过黄极与黄道垂直的大圆称黄经圈)为主圈,以春分点 ♈ 为主点.表示天体位置的两个球面



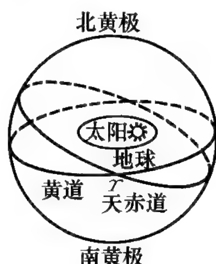
坐标为黄纬 β、黄经 λ.黄纬是自黄道沿通过天体的黄经圈到天体的弧段,自黄道向南黄极、北黄极分别计量,各从 0°到 90°,在黄道以北为正,在黄道以南

为负. 黄经是春分点和通过天体的黄经圈在黄道上的垂足间的弧段,自春分点沿黄道向夏至点方向,即沿逆时针方向(从北黄极看去)计量,从 0°到 360°. 天体的黄道坐标值不因周日视运动和观测地点的变化而变化.

黄经圈(longitude circle) 见“黄道坐标系”.

黄极(ecliptic pole) 见“黄道坐标系”.

黄道(ecliptic) 天球上的一个大圆. 指地球公转轨道平面与天球相交的大圆,即太阳周年视运动在天球上恒星之间的路径. 由于行星、月球的摄动,地球公转轨道平面(黄道面)在空间的位置不断变化,因此严格的说法是地球的瞬时平均轨道平面与天球相交的大圆. 黄道与天赤道的交角,称为黄赤交角. 其值变化于 22°59′和 24°36.2′间,每年减少 0.54″.



黄经(celestial longitude) 黄道坐标系的一个坐标. 指春分点与通过天体的黄经圈在黄道上的垂足间的弧段,或两者在北黄极处所成的两面角,以 λ 表示. 自春分点沿黄道朝夏至点方向或逆时针方向(自北黄极观看)计量,自 0°到 360°.

黄纬(celestial latitude) 黄道坐标系的一个坐标. 指黄道沿通过天体的黄经圈到天体的弧段,以 β 表示. 向南北黄极方向分别计量,各自从 0°到 90°,向北为正,向南为负.

银道坐标系(galactic coordinate system) 天文坐标系的一种. 以银道(银河带中心线与天球相交的大圆,与天赤道相交成 62.6°的交角)为基圈,以北银极(其赤道坐标为

$$\alpha = 12\text{h}49\text{min}, \quad \delta = +27°24',$$

历元 1950.0 年)为基点,以通过银河系中心 O(简称银心)方向的银经圈(通过银极与银道垂直的大圆)为主圈. 表示天体位置的银道坐标为银纬 B 和银经 L. 银纬 B 是自银道沿过天体的银经圈到天体的弧段,自银道向北银极方向为正,向南银极方向为负,各从 0°到 90°. 银经 L 是通过银心(其赤道坐标为



$$\alpha = 17\text{h}42.6\text{min}, \quad \delta = -28°57',$$

历元 1950.0 年)方向的银经圈与通过天体的银经圈在银道上所夹的弧段,自银心方向沿逆时针方向(在北银极处看)量度,从 0°到 360°(1958 年前采用的旧银道坐标系中,自银道与天赤道的升交点量起,银心的旧银道坐标为银经 327°41′,银纬 −1°24′).

银道(galactic equator) 天球上的一个大圆. 指银河系的平均平面与天球相交的大圆. 它和银河的中心线非常接近. 银道所在的平面称为银道面. 银道与天赤道相交于两点:升交点(银道自南向北通过天赤道的交点)和降交点(银道自北向南通过天赤道的交点),交角为 62°36′. 银道是天体银道坐标系的基本圈.

银极(galactic pole) 天球银道坐标系的基本点. 天球上距银道 90°的点称为银极. 在北的称北银极,在后发座内;在南的称南银极,在玉夫座内. 银极与天极的角距为 62°36′. 北银极是天体银道坐标系的基本点,其赤道坐标值为 12h49min = 192°15′;赤纬为 +27°24′(历元 1950.0 年),自 1958 年采用.

银经(galactic longitude) 天体银道坐标系的一个坐标. 指通过银心方向的银经圈与通过天体的银经圈在银道上所夹的弧段,以 L 表示. 自银心方向在银道上沿逆时针方向(自北银极处看去)量度,自 0°到 360°. 1958 年以前采用旧银道坐标系. 银经自银道与天赤道的升交点量起,在旧银道坐标系中,银心的旧银经为 327°41′,旧银纬为 −1°24′.

银纬(galactic latitude) 天体银道坐标系的一个坐标. 指自银道沿通过天体的银经圈到天体的弧段,以 B 表示. 自银道向南北银极方向分别计量,自 0°到 90°,向北银极方向为正,向南银极方向为负.

白道(moon's path) 天球上的一个大圆. 月球绕地球公转的轨道在天球上的投影称为白道. 它与黄道相交于两点. 月球在白道上从黄道以南运动到黄道以北的那个交点称为升交点,与此相对的另一交点称为降交点. 白道和黄道的交角在 4°57′～5°19′之间变化,平均值约为 5°9′,变化周期约为 173d. 由于太阳对月球的引力,使月球交点沿黄道向西移动,称为交点退行. 交点每年向西移动 19°21′,约 18.6a 完成一周,这就引起月球的赤纬变化. 当升交点与春分点重合时,黄道位于白道与赤道之间,白道与赤道的交角为黄赤交角与黄白交角之和;当降交点与春分点重合时,白道位于赤道与黄道之间,白道与赤道的交角为黄赤交角与黄白交角之差.

天文三角形(astronomical triangle) 一种球面三角形. 以观测者的天顶 Z、天极 p 和天体 σ 为顶点的球面三角形称为天文三角形. 即通过天体的地平经圈、赤经圈和天球子午圈三个大圆所组成的球面三角形. 应用球面三角学的公式可以导出天文三角形中联系天体位置和观测者位置的若干关系式,进行天体地平坐标和赤道坐标间的换算,并根据天体坐标计算观测者位置,在天体测量学和大地测量学中有广泛应用.

周日视运动(apparent diurnal motion)　一种天文现象.观测者随地球每日自西向东自转一周时看到天体在天球上所作的与地球自转方向相反的相对视运动称为周日视运动.天体作周日视运动时,沿着与天赤道平行的小圆称周日平行圈.周日视运动现象随观测者的地理纬度不同而异:以北半球观测点为例,在地理纬度为 φ 的地方,周日平行圈与地平圈成 $90°-\varphi$ 倾角.赤纬 $\delta \geqslant 90°-\varphi$ 的天体的周日平行圈,全部在地平圈上,天体永不下落,称为恒显圈.赤纬 $\delta \leqslant -(90°-\varphi)$ 的天体的周日平行圈,全部在地平圈下,天体永不升起称为恒隐圈.在赤道,天体周日平行圈垂直于地平圈,天体直升直落.在两极,周日平行圈平行于地平圈,没有升与落.天体在作周日视运动时,经过观测者的子午圈称为中天,经过观测者的地平圈称为出没,经过卯酉圈时称为过卯酉圈(仅对赤纬小于观测者地理纬度的天体).对于赤纬大于地理纬度的天体,周日视运动时其周日平行圈与地平经圈相切时称为大距.由于太阳、月亮的赤纬变化较快,其周日视运动的轨迹在每天都有显著的变化.

中天(culmination or meridian passage)　天文学的基本概念之一.天体在作周日视运动中两次通过观测者的子午圈的时刻或状态称为中天.离天顶较近的一次称上中天,离天顶较远的一次称下中天.天体上中天时,其地平高度达到最大值,其时角为零.天体下中天时,其地平高度达到最小值,时角为12h,对于地理纬度为 φ 的观测地点,在天顶以南上中天的天体(赤纬为 δ),其地平高度

$$h=90°-(\varphi-\delta),$$

方位角 $A=0°$;在天顶以北上中天时,

$$h=90°-(\delta-\varphi),$$

方位角 $A=180°$;下中天的天体的

$$h=(\varphi+\delta)-90°,$$

方位角 $A=180°$.天体上中天的恒星时 $s=\alpha$(α 为天体赤经);下中天的恒星时 $s=\alpha \pm 12\text{h}$.

大距(elongation)　天文学的基本概念之一.有下列两种不同的定义:

1. 天体在周日平行圈上运动,当周日平行圈与过该天体的地平经圈相切时称为大距.其赤纬大于地理纬度的天体才有大距.能发生大距的天体,在一恒星日内有两次大距,一次在东边称为东大距,一次在西边称为西大距.大距时天体的星位角(天体的赤经圈与地平经圈在天体处的夹角)$q=90°$,这时天体只有垂直运动,其方位角的变化率最大.

2. 在地球上观测内行星与太阳之间的角距最大值亦称大距,在太阳以东称为东大距,在太阳以西称为西大距.例如,水星的大距是 28°,金星的大距是 48°.

星位角(parallactic angle)　见“大距”.

大气折射(atmospheric refraction)　一种物理现象.旧称蒙气差.天体光线穿过地球大气层时受到的折射称为大气折射,是在入射点离开原来方向移向竖直线方向的偏移.大气折射对天体的方位角没有影响,而使天体的天顶距变小.设视天顶距为 Z',大气折射为 ρ,则真天顶距 $Z=Z'+\rho$.当天体在天顶时,大气折射为零,离天顶越远,大气折射越大.计算大气折射的近似公式为

$$\rho=60.2''\tan Z',$$

单位为(角)秒.对于小于 30°的天顶距,误差小于 0.1″.对于大于 60°的天顶距,上式无法应用.在实际应用中,不采用公式计算,而是利用平均大气折射表及其订正表查阅,即以视天顶距 Z' 为引数,由平均折射表查出平均大气折射值 ρ_0,再以气温 T、实际气压 P 和视天顶距 Z' 为引数,自平均大气折射订正表中查取 A, B, α,则大气折射 $\rho=\rho_0(1+\alpha A+B)$.

蒙气差(astronomical refraction)　即“大气折射”.

视差(parallax)　观测者在不同位置处看同一目标的方向之差.Parallax 一词源自希腊词 para(意为相比之下)和 allassein(意为产生差异),意为视线移动,近处物体与远处物体相比视线方向显出差异.中译为视差.在天文学上指观测者随地球自转、公转在不同位置观测同一天体的方向之差.定义为天体对观测者移动直线(基线)所张的角.天体对地球半径所张的角,称为周日视差 p.天体在地平时的周日视差最大,称为周日地平视差 p_0.1751～1753 年,法国拉卡伊(LaCaille, N. L. de)和拉朗德(Lalande, J. J. F. de)首次在近于同一地理经线上的柏林天文台和好望角天文台同时观测月球,测定出月球的周日地平视差.恒星对垂直于地心到恒星连线的地球公转轨道半径的张角称为周年视差 π.测定出周年视差,则应用三角学公式可以求出天体距离 d,即 $d=r/(\sin \pi)$,r 为日地平均距离,即地球轨道半长径,π 为恒星周年视差.当 d 用秒差距为单位,π 以(角)秒为单位时,上式可简化为 $d=1/\pi$.1837～1839 年,俄国斯特鲁维(Струве, В. Я)、德国贝塞尔(Bessel, F. W.)和英国亨德森(Henderson, T.)才分别测定出织女星、天鹅座 61 星和南门二(半人马座 α)三颗恒星的视差.天文学中常用周年视差表示恒星的距离.离太阳最近的恒星为半人马座比邻星,其周年视差为 0.756″.1952 年,出版了《恒星视差总表》,包含 6000 颗恒星的视差,其后又陆续精确测定了数百颗恒星的视差.

周年视差(annual parallax) 天文学的基本概念之一. 恒星对垂直于地球中心到恒星方向线的地球轨道平均半径 r 所张的角度 π 称为周年视差. 恒星距离 d 与周年视差 π 的关系为 $\sin\pi=r/d$. 当距离 d 以秒差距作单位,π 以(角)秒作单位时,$d=1/\pi$. 故天文学中常用恒星周年视差表示恒星距离. 恒星的位置常用日心赤道坐标表示. 由恒星的地心赤道坐标换算为日心赤道坐标,要加恒星的周年视差改正. 设恒星的地心赤道坐标为(α',δ'),日心赤道坐标为(α,δ),周年视差为 π,换算公式为

$$\begin{cases}(\alpha'-\alpha)=\pi(Yc-Xd),\\(\delta'-\delta)=\pi(Yc'-Xd'),\end{cases}$$

其中 c,d,c',d' 称为恒星常数,

$$c=\frac{1}{15}\cos\alpha\sec\delta,$$
$$c'=\tan\varepsilon\cos\delta-\sin\alpha\sin\delta,$$
$$d=\frac{1}{15}\sin\alpha\sec\delta,$$
$$d'=\cos\alpha\sin\delta,$$

(X,Y)为太阳的地心直角坐标.

周日视差(diurnal parallax) 天文学的基本概念之一. 天体对观测地点的地球半径所张的角度,即从地面上的观测者与从地心观测天体的方向之差称为周日视差. 它随天体的周日视运动而变化,天体在地平时,周日视差最大,称为周日地平视差;天体在中天时,周日视差最小;天体在天顶时,周日视差为零. 对于太阳系天体,测出其周日地平视差 p_0(如图). 利用三角学公式

天顶Z
σ_1
p_1
M
σ_2
p_0
R
d
O

$$d=\frac{R}{\sin p_0},$$

可求出天体到地球的距离 d,式中 R 为地球平均半径. 在同一子午线上相距很远的两个地点同时观测同一天体,可以测定出天体的周日地平视差. 对于太阳、月球和行星,需要将由地面观测所测定的赤道坐标加周日视差改正,换算为地心赤道坐标. 对于太阳和行星,换算公式为

$$\begin{cases}\alpha'-\alpha=-\rho p_0\cos\varphi'\sec\delta\sin(s-\alpha),\\ \delta'-\delta=p_0\big[-\rho\sin\varphi'\cos\delta+\rho\cos\varphi'\sin\delta\cos(s-\alpha)\\ \qquad+\frac{1}{2}(\alpha'-\alpha)\rho\sin\delta\sin(s-\alpha)\big].\end{cases}$$

对于月球,相应公式为

$$\begin{cases}\tan(\alpha'-\alpha)=\dfrac{-\rho\cos\varphi'\sin p_0\sec\delta\sin(s-\alpha)}{1-\rho\cos\varphi'\sin p_0\sec\delta\cos(s-\alpha)},\\ \tan(\delta'-\delta)=\dfrac{-\rho\beta\sin p_0\sin(\gamma-\delta)}{1-\rho\beta\sin p_0\cos(\gamma-\delta)},\end{cases}$$

式中(α',δ')为地面赤道坐标;(α,δ)为地心赤道坐标;ρ 为地心到地面观测地点的向径;s 为观测瞬间的地方恒星时;p_0 为周日地平视差;φ' 为地心纬度;β,γ 为辅助量,由下式决定:

$$\begin{cases}\beta\sin\gamma=\sin\varphi',\\ \beta\cos\gamma=\cos\varphi'\dfrac{\cos\left[s-\frac{1}{2}(\alpha'+\alpha)\right]}{\cos\frac{1}{2}(\alpha'+\alpha)}.\end{cases}$$

光行差(aberration) 天文学的基本概念之一. 指运动着的观测者所看到的天体的视方向与同一时刻静止的观测者所看到的天体的真方向之间的偏差. 由于光速有限,观测者随着地球作自转和公转,因此观测者所看到的天体的方向是观测者的速度和天体光速的合成方向. 此合成方向与天体真实方向之差即光行差. 由地球公转产生的光行差称为周年光行差;由地球自转产生的光行差称为周日光行差. 观测者的运动速度与光速之比称为光行差常数 K,是天文学的导出常数之一. 1976 年,国际天文学联合会决定,自 1984 年起采用 $K=20.495\,52''$(历元 2000.0 年). 在周年光行差影响下,天体向着观测者运动方向与天球的交点(奔赴点)的方向位移,周日光行差的影响使天体向东点位移.

周年光行差(annual aberration) 一种光行差. 观测者随地球在空间作绕太阳的周年运动所引起的光行差称为周年光行差. 在周年光行差影响下,天体向着地球周年运动奔赴点方向运动. 周年光行差对天体的赤道坐标影响为

$$\alpha'-\alpha=-K\sin\lambda_\odot\sin\alpha\sec\delta$$
$$\qquad-K\cos\lambda_\odot\cos\varepsilon\cos\alpha\sec\delta,$$
$$\delta'-\delta=-K\sin\lambda_\odot\cos\alpha\sin\delta$$
$$\qquad-K\cos\lambda_\odot\cos\varepsilon(\tan\varepsilon\cos\delta$$
$$\qquad-\sin\alpha\sin\delta),$$

式中(α,δ)为真实位置;(α',δ')为视位置;$\lambda_\odot$ 为太阳黄经,ε 为黄赤交角. 若令

$$C=-K\cos\varepsilon\cos\lambda_\odot,\qquad D=-K\sin\lambda_\odot,$$
$$c=\frac{1}{15}\cos\alpha\sec\delta,\qquad d=\frac{1}{15}\sin\alpha\sec\delta,$$
$$c'=\tan\varepsilon\cos\delta-\sin\alpha\sin\delta,\qquad d'=\cos\alpha\sin\delta,$$

则 $\alpha'-\alpha=Cc+Dd,\ \delta'-\delta=Cc'+Dd'$.

或令 $C=h\sin H,\ D=h\cos H,\ i=C\tan\varepsilon$,则

$$\alpha'-\alpha=h\sin(H+d)\sec\delta,$$
$$\delta'-\delta=h\cos(H+d)\sin\delta+i\cos\delta.$$

式中 C,D 称为贝塞尔日数;c,d,c' 和 d' 称为恒星常数;i,h 和 H 称为独立日数,可从天文年历查出.

周日光行差(diurnal aberration) 光行差的一种. 观测者随地球作周日运动而引起的光行差称为周日光行差. 地球赤道上的周日光行差为

$$K'_0=\frac{v}{c}=0.32'',$$

式中 v 为地球自转引起的观测者运动速度,c 为光速. 任一地理纬度 φ 处的周日光行差常数 $K'_{\varphi}=0.32''\cos\varphi$,周日光行差的奔赴点是东点. 周日光行差对天体坐标的影响为

$$\alpha'-\alpha=0.021^{s}\cos\varphi\sec\delta\cos t,$$
$$\delta'-\delta=0.32''\cos\varphi\sin\delta\sin t,$$

式中(α',δ')为天体的视坐标;(α,δ)为顾及周日光行差的天体视坐标. 由上两式可知,在子午圈上,周日光行差对赤经影响最大,在天体时角 $t=6\text{h}$ 或 $t=18\text{h}$ 时,天体赤经完全不受影响. 当 $t=6\text{h}$ 或 18h,周日光行差对赤纬影响最大,而在子午圈上则不受影响. 周日光行差对恒星天顶距和地平经度的影响为

$$Z-Z'=0.32''\cos\varphi\sin A\cos Z,$$
$$A-A'=0.32''\cos\varphi\cos A\csc Z.$$

周日光行差影响很小,一般对赤纬小于 80°的星可不需顾及.

岁差(precession) 天文学的基本概念之一. 指由于春分点沿黄道缓慢西移(每年约 50.2″),而使回归年比恒星年短的现象. 是古希腊天文学家喜帕恰斯(Hipparchus,(R.))于公元前 150 年前后首先发现的. 产生的原因是:日、月、行星对地球赤道凸出部质量的吸引,造成地轴(天轴)的进动,即地轴绕通过黄极的轴线按顺时针方向旋转,造成平天极绕黄极沿着一个小圆在约 26000a 中顺时针方向旋转一周,从而使天赤道、春分点位置发生变化. 日、月引力造成的岁差称为日月岁差;行星引力造成的岁差称为行星岁差;二者合称为总岁差. 岁差使天体在天球上的平位置发生改变. 主要分量是沿黄经方向每年约增加 50.2″. 在不同历元,春分点位置不同,同一恒星坐标值也不同;需把恒星位置化为属于所需的某一春分点的位置,即加上岁差改正. 设(α_0,δ_0)为天体相对于 t_0 时的平赤道坐标,(α_1,δ_1)为相对于 t_1 时的平赤道坐标,则岁差改正为

$$\begin{cases}\alpha_1-\alpha_0=(m+n\sin\alpha_0\tan\delta_0)\tau,\\ \delta_1-\delta_0=n\cos\alpha_0\tau,\end{cases}$$

式中 $\tau=t_1-t_0$,$m=\psi'\cos\varepsilon-\lambda'$,$n=\psi'\sin\varepsilon$,$\psi'$ 为日月岁差造成的平春分点在黄道上的运动速度,$\psi'=50.3708''+0.0050''T$,$\lambda'$ 为行星岁差造成的平春分点在赤道上的运动速度,$\lambda'=0.1247''-0.0188''T$. T 为从 1900.0 年起算的回归世纪数. 设某一瞬间的儒略日为 JD,则对应的 T 为

$$T=\frac{\text{JD}-2415020.3135}{36524.22}.$$

当 τ 间隔较长时,则需计及岁差改正的二次项及三次项.

总岁差(general precession) 见"岁差".

章动(nutation) 天文学的基本概念之一. 地轴进动时伴随的许多短周期的摆动称为章动,其中最主要的摆动周期是 18.6a. 章动使天体坐标发生微小变化. 设某一瞬间天体的平赤道坐标为(α_0,δ_0),真赤道坐标为(α,δ),章动对天体黄经的影响为 $\Delta\psi$,对黄赤交角的影响为 $\Delta\varepsilon$,对天体黄纬没有影响,则

$$\begin{cases}\alpha-\alpha_0=\Delta\psi(\cos\varepsilon+\sin\varepsilon\sin\alpha\tan\delta)\\ \qquad\qquad -\Delta\varepsilon\tan\delta\cos\alpha,\\ \delta-\delta_0=\Delta\psi\sin\varepsilon\cos\alpha+\Delta\varepsilon\sin\alpha,\end{cases}$$
$$\begin{aligned}\Delta\psi=&-(17.2327''+0.01737'')\sin\Omega\\ &+(0.2088''+0.00002''T)\sin 2\Omega\\ &-(1.2729''+0.00013''T)\sin 2L+\cdots,\\ \Delta\varepsilon=&-(9.210''+0.00091''T)\cos\Omega\\ &-(0.0904''-0.00004''T)\cos 2\Omega\\ &+(0.5522''-0.00029''T)\cos 2L+\cdots,\end{aligned}$$

式中 Ω 为白道升交点的平黄经,自瞬时平春分点起算,L 为太阳几何平黄经,T 为自 1990 年 1 月 0.5 日历书时起算的儒略世纪数.

天体位置(celestial position) 天体的天球坐标. 用天文仪器直接测定的天体的天球坐标,并已消除仪器误差的影响,称为天体的观测位置. 观测位置作了大气折射、周日视差和周日光行差改正后的天体地心天球坐标,称为天体视位置,又称天体视坐标,是相对观测瞬间的真赤道和真春分点的. 视位置作了周年视差和周年光行差改正后得到的日心天球坐标称天体真位置,是相对于观测瞬间的真赤道和真春分点的. 真位置作了章动改正后得到的日心天球坐标,称为天体平位置,是相对于观测瞬间的平赤道和平春分点的. 平位置减去岁差和自行所得的天球坐标是岁首平位置. 星表、星图给出的天体坐标即为某一历元的岁首平位置. 它们之间的关系是

视位置=(岁首平位置+岁差+自行)+章动
　　+周年视差+周年光行差
　=(平位置+章动)+周年视差
　　+周年光行差
　=真位置+周年视差+周年光行差.

天体观测位置(celestial observational position) 见"天体位置".

天体视位置(celestial apparent position) 天体位置的一种. 由直接观测位置作了大气折射、周日光行差和周日视差改正后,所得天体的地心天球坐标称为天体视位置. 它相当于一个位于地心的观测者在没有大气的情况下所见天体的位置. 坐标为观测瞬间的真赤道、真春分点. 天体视位置受到光行差和视差、岁差、章动的影响,对于恒星,仅少数近距恒星要作周年视差改正. 某些恒星视位置可从天文年历查出,有些恒星的视位置需要从星表中给出的星表历元平位置算出. 设(α_0,δ_0)为恒星在星表历元 t_0 时的平位置;(α',δ')为在观测年首 t'时的平位置;

(α,δ)为观测瞬间恒星视位置. 那么，首先由下式计算出年首平位置：

$$\begin{cases}\alpha'=\alpha_0+(t'-t_0)AV_\alpha+\left(\dfrac{t'-t_0}{200}\right)^2SV_\alpha+\left(\dfrac{t'-t_0}{100}\right)^3 3^{\mathrm{d}}_{\alpha}t,\\ \delta'=\delta_0+(t'-t_0)AV_\delta+\left(\dfrac{t'-t_0}{200}\right)^2SV_\delta+\left(\dfrac{t'-t_0}{100}\right)^3 3^{\mathrm{d}}_{\delta}t.\end{cases}$$

AV_α, AV_δ 为周年岁差

$$\left(\frac{\mathrm{d}\alpha}{\mathrm{d}t},\frac{\mathrm{d}\delta}{\mathrm{d}t}\right)$$

和一年内恒星的自行(μ_α,μ_δ)之和，为恒星位置周年变化量，简称年变.

$$SV_\alpha=100\times\frac{\mathrm{d}^2\alpha}{\mathrm{d}t^2},\quad SV_\delta=100\times\frac{\mathrm{d}^2\delta}{\mathrm{d}t^2},$$

为 100 年内周年岁差的变化，称为百年变化.

$$3^{\mathrm{d}}_{\alpha}t=\frac{100^3}{6}\frac{\mathrm{d}^3\alpha}{\mathrm{d}t^3},\quad 3^{\mathrm{d}}_{\delta}t=\frac{100^3}{6}\frac{\mathrm{d}^3\delta}{\mathrm{d}t^3},$$

称为第三项，均可由星表中查出. 其次，由下式算出观测瞬间视位置(未计算周年视差改正)：

$$\begin{cases}\alpha=\alpha'+(A+A')a+(B+B')b\\ \qquad +Cc+Dd+E+\mu_\alpha\tau,\\ \delta=\delta'+(A+A')a'+(B+B')b'\\ \qquad +Cc'+Dd'+\mu_\delta\tau.\end{cases}$$

式中$(A+A')a+(B+B')b+E$，$(A+A')a'+(B+B')b'$为岁差和章动改正，$Cc+Dd$，$Cc'+Dd'$为光行差改正；A,B,C,D,E,A',B'称为贝塞尔日数，可由天文年历查出；a,b,c,d,a',b',c',d'称为恒星常数，只与恒星坐标有关，可在某些星表查出；μ_α,μ_δ为恒星赤经周年自行和赤纬周年自行；τ为观测瞬间 t 与年首 t'之时间差 $\tau=t-t'$，t 在上半年，t'取当年年首，t 在下半年，t'取下一年年首，使 τ 不超过半年. 另一种视位置的计算公式为

$$\begin{cases}\alpha=\alpha'+f+g\sin(G+\alpha')\tan\delta'\\ \qquad +h\sin(H+\alpha')\sec\delta'+\mu_\alpha\tau,\\ \delta=\delta'+g\cos(G+\alpha')\\ \qquad +h\cos(H+\alpha')\sin\delta'+i\cos\delta'+\mu_\delta\tau.\end{cases}$$

式中 f,g,h,i,G,H 称为独立日数，可由天文年历查出.

天体真位置(celestial true position)　天体位置的一种. 指天体视位置消除了周年视差和光行差后得到的天体的日心天球坐标. 严格说来，应是天体相对于太阳系质心的天球坐标. 它相当于一个位于日心(太阳系质心)的观测者所见天体的位置，坐标系为观测瞬间的真赤道和真春分点. 设观测瞬间天体视位置为(α,δ)，观测瞬间对应的天体真位置为$(\alpha_{真},\delta_{真})$，则

$$\begin{cases}\alpha_{真}=\alpha-(Cc+Dd)-\pi(Yc-Xd),\\ \delta_{真}=\delta-(Cc'+Dd')-\pi(Yc'-Xd'),\end{cases}$$

式中

$$c=\frac{1}{15}\cos\alpha\sec\delta,\qquad c'=\tan\varepsilon\cos\delta-\sin\alpha\sin\delta,$$

$$d=\frac{1}{15}\sin\alpha\sec\delta,\qquad d'=\cos\alpha\sin\delta,$$

$$C=-K\cos\lambda_{\odot}\cos\varepsilon,\quad D=-K\sin\lambda_{\odot}.$$

$\lambda_{\odot}$为太阳黄经；ε 为黄赤交角；$(Cc+Dd)$，$(Cc'+Dd')$为光行差改正；$\pi(Yc-Xd)$，$\pi(Yc'-Xd')$为周年视差改正；π 为周年视差；(X,Y)为太阳的地心直角坐标；C,D 称为贝塞尔日数，可从天文年历查出；c,c',d,d'称为恒星常数，可从专用历表中查出.

天体平位置(celestial mean position)　天体位置的一种. 以某一瞬间的平赤道和平春分点为参考的天体坐标称为天体平位置. 它相当于一个位于日心(太阳系质心)的观测者所见的天体位置. 平位置是真位置消除章动影响后的坐标. 设$(\alpha_{真},\delta_{真})$为观测瞬时 t 的天体真坐标，$(\alpha_{平},\delta_{平})$为天体的平坐标，(α_0,δ_0)为年首 t_0 的平坐标，则

观测瞬时平坐标＝岁首平坐标＋岁差＋恒星自行，

即

$$\alpha_{平}=\alpha_0+(m+n\sin\alpha_0\tan\delta_0)\tau+\mu_\alpha\tau,$$

$$\delta_{平}=\delta_0+n\cos\alpha_0\tau+\mu_\delta\tau,$$

$$m=\psi'\cos\varepsilon-\lambda',$$

$$n=\psi'\sin\varepsilon,$$

μ_α,μ_δ 为恒星周年自行，$\tau=t-t_0$，ψ'为日月岁差造成的平春分点在黄道上的运动速度，

$$\psi'=50.3708''+0.0050''T,$$

λ'为行星岁差造成的平春分点在赤道上的运动速度，

$$\lambda'=0.1247''-0.0188''T,$$

T 为从 1900.0 年起算的回归世纪数.

观测瞬时平坐标＝观测瞬时真坐标－章动，

即

$$\alpha_{平}=\alpha_{真}-(\Delta\psi+d\psi)(\cos\varepsilon+\sin\varepsilon\sin\alpha_0\tan\delta_0)$$
$$\quad-(\Delta\varepsilon+d\varepsilon)\cos\alpha_0\tan\delta_0,$$
$$\delta_{平}=\delta_{真}-(\Delta\psi+d\psi)\sin\varepsilon\cos\alpha_0$$
$$\quad+(\Delta\varepsilon+d\varepsilon)\sin\alpha_0,$$

$\Delta\psi,\Delta\varepsilon$ 表示章动长期项，$d\psi,d\varepsilon$ 表示章动短期项，ε 为黄赤交角.

基本天体测量学(fundamental astrometry)　天体测量学的一个分支学科. 主要研究内容为：精确测定恒星位置、恒星自行和岁差常数，最后编制成基本星表. 它的任务是通过基本星表建立起一个基本的天文参考坐标系. 编制基本星表要将各初始星表归算到同一历元，并且归算到同一天文常数系统. 初始星表中的星位存在下列误差：

1. 由于观测本身的随机误差经传播而产生的随机误差.

2. 由于观测中的系统误差而产生的系统误差.

3. 由于在统计平差中求得的归算参数存在不可避免的随机误差而产生的系统误差.

因此,在综合各初始星表的资料时,应在恒星星位中尽量减小星表的系统误差和随机误差.把全部初始星表都归算到同一坐标系统——基本星表系统.为了保证在较长时间内能复制出坐标系并保持其精度不变,还要推算出恒星自行的估计值.一本基本星表所提供的天文参考坐标系,一般只能用20年左右.必须不断地利用更多的最新观测成果来更新基本星表的坐标系统.现在建立基本星表只需改进现有的基本系统,即求出现有基本星表的改正值:星位改正值$\Delta\alpha$和$\Delta\delta$、系统定向改正值$\Delta\alpha_0$和$\Delta\delta_0$以及系统自行改正值$\Delta\mu_\alpha$和$\Delta\mu_\delta$.

惯性参考系(inertial frame of reference) 一种参考系.牛顿运动定律在其中能准确成立的参考系称为惯性参考系.相对惯性参考系作匀速直线运动的系统,也是惯性参考系.物体在各个惯性参考系内,都服从同样的运动规律.研究建立惯性参考系的原则和实施方法,是天体测量学的一个基本任务.在天体测量学中,惯性参考系是以基本星表的形式来近似实现的.根据国际天文学联合会决议,自1984年起,以第五基本星表(FK5)中的约5000颗均匀分布于全天球、视星等亮于9等的恒星的星位和自行系统作为实际使用的惯性参考系.由于FK5星表以赤道面和春分点为参考面和参考点,这就涉及太阳系模型或者地球模型,要受到太阳系甚至银河系运动的影响,因此这只是一种近似惯性参考系.随着甚长基线干涉测量技术的发展和观测精度的提高,目前正在建立以河外射电源为参考的参考系.这个参考系的建立,不涉及黄道、赤道和其他动力学假设.河外射电源的距离遥远,自行可以忽略.相对于这些遥远参考目标来说,参考系坐标轴的方向是固定的,所以在描述物体运动的微分方程中,并不含有参考系的转动带来的虚假项.因而这个参考系是比较理想的惯性参考系,建立这种参考系的惟一假定,是宇宙膨胀中不存在横向成分.这样的惯性参考系用一个河外源星表来具体体现.

初始星表(primary catalogue) 刊载恒星在某一历元的平赤道坐标的表册.它是依据某架仪器在不太长期间内的观测所得到的恒星坐标编成的.每颗恒星在观测期间内都观测多次,以减弱随机误差对星位的影响.在不同时刻所得到的恒星视位置都归算到某一个贝塞尔岁首,要计算岁差、章动、光行差和视差的影响,归算到岁首的平位置,习惯上采用公历纪元数的25的整倍数的某年岁首作为星表的历元.1984年后统一采用2000.0作为星表历元.按所采用的观测方法不同,可分为绝对星表和相对星表.前者的恒星位置的测定完全不依赖于过去已知的数据,而用绝对测定方法测定星位,除观测恒星外,还必须进行坐标原点的测定并独立确定仪器误差.绝对星表包含星数较少,通常只有数千颗恒星.后者的恒星位置是由相对测定得出的,即将大量确定位置的恒星跟少量已按绝对测定方法确定出位置的恒星一起进行观测求出两者的位置差,再根据已知恒星的位置求出被测星的位置.初始星表包含恒星的重要原始资料,可用于编制基本星表,建立精确的基本坐标系.根据若干天文台不同时期的几本初始星表的位置,综合计算而成的星表称为编制星表.

绝对星表(absolute catalogue) 见"初始星表".

相对星表(relative catalogue) 见"初始星表".

基本星表(fundamental catalogue) 一种星表.它是将若干天文台不同时期的数本初始绝对星表以及一些精度较高的相对星表综合处理后得到的高精度星表.初始星表中的星位存在由于观测本身的随机误差传播而产生的随机误差和由于观测中的系统误差而产生的系统误差以及在统计平差中求归算参数过程存在不可避免的随机误差而产生的系统误差,因此,在综合处理初始星表资料时,应在恒星星位中尽量减小星表的系统误差和随机误差.基本星表是一切星表的基础,主要用作天文坐标参考系和恒星位置相对测定时的定标星系统.现代主要的基本星表有以下几种:FK4星表,由德国海德堡天文计算研究所弗里克(Fricke,W.)和柯弗(Kopff,A.)编纂,于1963年出版,刊载1535颗全天分布的恒星位置和自行;FK5星表,1984年出版,刊载5000颗全天分布的视星等亮于9等的恒星的位置和自行,它是目前国际广泛使用的基本星表,作为实际使用的惯性坐标系;GC星表由博斯(Boss,B.)编纂,于1937年出版,包含有33324颗恒星的位置和自行;N30星表由摩根(Morgan,H. R.)于1952年编纂出版,共有5268颗恒星的基本星表.

天文常数系统(system of astronomical constants) 天文常数组成的数组.指用于编算天文年历和计算天体位置的一组与地球有关的常数(地球大小、形状、重力场,地球的轨道要素,岁差常数,章动常数和光行差常数)和太阳、月球及行星的质量等数值,是研究天体的空间分布和运动规律的基本数据.天文常数系统的建立是天体测量学的重要研究内容之一.各天文常数中,有些天文常数相互之间受着物理定律的制约,各天文常数应该严格符合理论推出的关系式.此外,各常数的采用值与最精确的测定值之差应在测量精度范围之内.1896年,在巴黎召开的国际基本恒星会议和1911年举行的国际天文年历会议曾经一致决定,统一采用美国天文学家纽科姆(Newcomb,S.)所推荐的一组天文常数,包

括14个天文常数,称为纽科姆天文常数系统.该天文常数系统从1896年至1967年底共使用了70余年.1964年,国际天文学联合会第12届大会通过了“1964年国际天文学联合会天文常数系统”,从1968年起采用,包含23个天文常数及一组大行星质量的采用值,使用到1983年底.1976年,国际天文学联合会通过了目前使用的“1976年国际天文学联合会天文常数系统”,从1984年起采用,包含19个天文常数和一组大行星质量采用值.有:

1个定义常数:高斯引力常数

$$k=0.01720209895;$$

9个基础常数:光速 $c=299792458\text{m/s}$,每天文单位距离的光行时 $\tau_A=499.004782\text{s}$,地球赤道半径 $a_e=6378140\text{m}$,地球动力学形状因子 $J_2=0.00108263$,地心引力常数 $GE=3.986005\times10^{14}\ \text{m}^3/\text{s}^2$,引力常数 $G=6.672\times10^{-11}\ \text{m}^3/(\text{kg}\cdot\text{s}^2)$,月地质量比 $\mu=0.01230002$,黄经总岁差 $(J2000.0)p=5029''.0966$(每世纪),黄赤交角

$$(J2000.0)\varepsilon=23°26'21.448'');$$

9个导出常数:章动常数 $(J2000.0)N=9.2025''$,天文单位距离 $A=c\tau_A=1.49597870\times10^{11}\text{m}$,太阳视差 $\pi_\odot=\arcsin(a_e/A)=8.794148''$,光行差常数 $(J2000.0)\kappa=20.49552''$,新的标准历元为公元2000年1月15日,即儒略日2451545.0,记为 $J2000.0$.地球扁率 $f=0.00335281=1/298.257$,日心引力常数 $GS=A^3k^2/D^2=1.32712438\times10^{20}\ \text{m}^3/\text{s}^2$,日地质量比 $S/E=GS/GE=332946.0$,日与地月系质量比 $(S/E)/(1+\mu)=328900.5$,太阳质量 $S=GS/G=1.9891\times10^{33}\text{kg}$;

行星质量倒数:水星为6023600,金星为408523.5,地月系为328900.5,火星为3098710,木星为1047.355,土星为3498.5,天王星为22869,海王星为19314,冥王星为3000000.

黄经总岁差(general precession in longitude) 亦称岁差常数.天文常数系统的基础常数之一.是在日月岁差和行星岁差综合影响下平春分点在一回归世纪内沿黄道的岁差值,以 p 表示.以 p_1 表示日月岁差,λ 表示行星岁差,ε 表示黄赤交角,则有

$$p=p_1-\lambda\cos\varepsilon.$$

p_1 可根据大量恒星的观测资料来确定,λ 可由天体力学理论计算出来.1976年,国际天文学联合会第16届大会上通过,将美国天文学家纽科姆(Newcomb, S.)得出的对应1900.0历元旧值 $p=5025.64''$ 归算到历元2000.0,并加银河系自转的改正,又在计算行星岁差时采用新的行星质量数据所得到的新值 $p=5029.0966''$.

岁差常数(precession coefficient) 即“黄经总岁差”.

黄赤交角(obliquity of ecliptic) 天文常数系统的基础常数之一.黄道与天赤道的交角称为黄赤交角,以 ε 表示.黄道面位置的改变,使黄赤交角有长期变化.1901年,美国天文学家纽科姆(Newcomb,S.)提出的计算黄赤交角的公式为

$$\varepsilon=23°27'8.26''-46.845''T-0.0059''T^2+0.00181''T^3,$$

式中 T 是1900年起算的儒略世纪数(一个儒略世纪等于36525日).当前每世纪减小47″左右,这种减小还会持续15000年左右,然后,黄赤交角转为增大.1976年,国际天文学联合会第16届大会决议,黄赤交角采用 $\varepsilon=23°26'21.448''$(历元2000.0年).

章动常数(nutational constant) 天文常数系统中的基础常数之一.章动椭圆的半长径指向黄极方向,半短径指向春分点方向,章动椭圆的半长径即为章动常数,用符号 N 表示.1896年,国际基本恒星会议决定采用美国天文学家纽科姆(Newcomb, S.)求得的 $N=9.210''$,沿用到1983年.1976年第16届国际天文学联合会通过 $N=9.2109''$(历元2000.0年).1979年第17届国际天文学联合会通过 $N=9.2025''$(历元2000.0年),1984年开始正式采用.新的章动常数不采用观测值,而由月地质量比、日月岁差和章动常数间的理论关系,先求出刚体地球的章动常数,然后换算成非刚体地球的数值,因此新章动常数作为导出常数.

天文单位[距离](astronomical unit [distance]) 天文常数系统的导出常数之一.指天文学中测量距离,特别是测量太阳系天体之间的距离的基本单位.1938年以前,天文单位[距离](A)是指在没有大行星摄动作用下,从地月系质心到日心的平均距离,或地月系质心绕日公转的无摄动椭圆轨道的半长径.根据开普勒第三定律,有

$$k=\frac{2\pi A^{3/2}}{T\ \sqrt{S+E+M}},$$

k 作为高斯引力常数,S,E,M 分别为太阳、地球和月球的质量,T 为地球公转周期.取 $A=1,S=1,E+M=0$,k 取高斯(Gauss,C. F.)所计算的值 $k=0.01720209895$,求出

$$T=365.2568983263\ \text{平太阳日}.$$

1938年,国际天文学联合会第6届大会将天文单位[距离]定义为:当公转周期为

$$T=365.2568983263\ \text{平太阳日时},$$

一个假想的零质量的无摄动行星的椭圆轨道的半长径.根据准确的 T 值和 S,E 值,可计算出地月系绕日公转轨道的半长径为1.00000003天文单位.由于地球运动受到其他行星摄动的影响,日地平均距

离实际为 1.000000236 天文单位[距离]. A 由光速 c 和天文单位距离光行时 τ_A 导出，即 $A=c\tau_A$. 1976 年，国际天文学联合会天文常数系统取

$$A=1.49597870\times10^{11}\text{m}.$$

从 1984 年起采用.

天文单位距离的光行时(light-time for astronomical unit distance) 天文常数系统的基础常数之一. 指光(电磁波)经过 1 天文单位[距离](A)所需要的时间 τ_A，即 $c\tau_A=A$，c 为光速. 20 世纪 60 年代以前，无法直接测定 τ_A，1961 年后，利用对行星雷达测距，并按天体力学理论精确推算出行星对地球的距离与日地平均距离之比，因此可以直接得到 τ_A. 根据美国麻省理工学院和美国喷气推进实验室在 1970～1974 年间的 4 次行星雷达测距的结果综合分析得到 $\tau_A=499.004782$ s. 1976 年，国际天文学联合会把它作为基础常数，从 1984 年起采用.

太阳视差(solar parallax) 天文常数系统的导出常数之一. 以 $\pi_\odot$ 表示. 确切名称是太阳赤道地平视差. 定义为

$$\sin\pi_\odot=\frac{a_e}{A},$$

a_e 为地球赤道半径，A 为天文单位[距离]. 20 世纪 60 年代以前，天文单位[距离]是通过测定太阳视差来推算的，由于天文单位[距离]是量度太阳系天体之间距离的基本单位，又是测定恒星三角视差的基线，因此测定太阳视差是天体测量学的重要问题. 现代的最精确观测是采用雷达天文方法，测定出天文单位距离的光行时 τ_A，由 $A=c\tau_A$ 求出天文单位[距离]A，c 为光速. 再从 A 导出 $\pi_\odot$. 在 1976 年国际天文学联合会天文常数系统中，$\pi_\odot=8.794148''$，从 1984 年起采用.

光行差常数(aberration constant) 天文常数系统中的导出常数之一. 地球公转的平均线速度 v 与光速 c 之比值称为光行差常数，以符号 κ 表示. 其关系式为

$$\kappa=\frac{v}{c},$$

以[角]秒为单位. 天文常数中的光行差常数指周年光行差常数. 首先由英国天文学家布拉德雷(Bradley, J.)于 1728 年得出 $\kappa=20.5''$；1842 年，俄国斯特鲁维(Струве, В. Я.)给出 $\kappa=20.4451''$，在全世界使用了将近半个世纪；1896 年，美国天文学家纽科姆(Newcomb, S.)得出 $\kappa=20.47''$，使用了将近 70a；1968 年，根据国际天文学联合会于 1964 年的决定，采用 $\kappa=20.4958''$. 自 1984 年起采用国际天文学联合会 1976 年通过的 $\kappa=20.49552''$(历元 2000.0 年)，是由公式

$$\kappa=\frac{Fk\tau_A}{86400}$$

计算并换算到标准历元 2000.0，式中 $F=1.0001413$，是一个辅助常数，它表示地球的平均速度与一假想行星公转速度的比值，k 为高斯引力常数，τ_A 为天文单位距离的光行时.

方位天文学(positional astronomy) 天体测量学的一个分支. 主要任务是：研究天体的方位和运动，建立惯性参考坐标系，测定恒星的视差和自行、双星和聚星系统的相对位置和位置变化. 为解决这些任务，按技术手段的历史发展顺序，曾采用目视观测、照相观测、射电干涉技术和近些年才发展起来的地球外层的空间观测技术. 特别是空间观测技术已摆脱地球大气的影响，它可用于改进星表坐标系统，精测自行和视差以及天体间的相对位置，远非地面观测可比拟. 这些手段既有分工又互相补充、互相促进，从而丰富和推动方位天文学内容的扩充和精度的提高，为解决天体演化、银河系结构和宇宙学新理论等课题迈向新阶段.

照相天体测量学(photographic astrometry) 天体测量学的一个分支. 主要任务是：天体相对位置的测定(包括对象有行星、彗星、流星和人造卫星等)；恒星自行和三角视差的测定；目视双星和聚星系统的相对位置和运动的测定；照相星表的编制；日全食时相对论效应的验证；彗尾等离子体在太阳风作用下出现的断尾和节点的运动的测定等. 照相天体测量学虽仅有一百多年的历史，但却极大地促进了恒星天文学和银河系结构的研究以及太阳系天体和人造卫星运动理论的发展，目前仍在不断更新技术手段，例如 CCD(电荷耦合器件)技术，又向观测暗星迈进一大步，同时由于空间技术的广泛应用，也给照相天体测量学带来活力和新的方向.

照相天体测量方法(method of photographic astrometry) 测量天体位置和运动的一种方法. 用照相方法测量天体位置和运动称为照相天体测量方法. 与目视观测相比，其优点是：在底片的视场范围内，能同时拍摄到大量的天体信息；由于照相底片对星光有累积作用，若适当延长曝光时间，可获得更暗的天体；底片又可长期保存，具有文献性. 照相定位的四个基本过程是：

1. 拍摄底片. 将底片盒固连在天体照相仪的焦平面上，为获取清晰星像，要求望远镜能准确跟踪天体的周日运动，为此需用导星镜，以确保跟踪精度.

2. 测量底片. 用坐标量度仪测量定标星和目标天体的平面直角坐标(X,Y)，可量度到 0.1 μm 的量级.

3. 解算底片常数. 首先对位置已知和均匀分布的定标星，计算它们在天球切平面上的理想坐标(ξ,

η)，借助联系公式

$$\xi = aX + bY + c,$$
$$\eta = dX + eY + f,$$

解算出底片联系常数 a,b,c,d,e,f，一般只需3颗定标星.更多的选取定标星，由最小二乘法解算是有利的.

4. 确定目标天体位置.通过已求出的底片常数 a,b,c,d,e,f，计算目标的 (ξ,η)，再变换成球面坐标，要求天体定位精度为0.1″时，需要天体照相仪有足够长的焦距，准确的星表和数量足够多又均匀分布的定标星.

天体照相仪(astrograph)　一种天文望远镜.专为天体照相设计的望远镜.有折射、反射和折反射望远镜三种，通常指折射望远镜.按用途和相对口径大致分为三类：

1. 长焦天体照相仪.焦距在10m以上，相对口径约为1/20～1/10，视场约2°，常用来测定恒星自行、三角视差以及目视双星的相对运动等.

2. 中焦天体照相仪.焦距在3～7m，相对口径约1/10～1/7，视场约3°左右，常用在小行星、彗星等精密定位.

以上两种望远镜均为两片型消色差物镜，并未作像散和场曲的改正.

3. 短焦天体照相仪.焦距短于4m，相对口径大于1/7，视场约5°左右，用于流星和人造卫星等的定位和照相测光工作，镜片为3片或更多，若要求视场在10°以上，则需折反射望远镜.

理想坐标(ideal coordinate)　照相天体测量学用语.是为确定人造卫星、小行星和彗星等天体位置而采用的一种过渡性坐标系.即将天体球面位置投影到天球切平面上的直角坐标系.照相底片平面与这个天球切平面平行.直角坐标系的 ξ 轴和 η 轴分别平行于赤纬圈和赤经圈.以底片光心 (α_0,δ_0) 为原点，定标星的理想坐标 (ξ,η) 通过下式可由其球面坐标 (α,δ) 求出：

$$\xi = \frac{\cot\delta\sin(\alpha-\alpha_0)}{\sin\delta_0+\cot\delta\cos\delta_0\cos(\alpha-\alpha_0)},$$
$$\eta = \frac{\cos\delta_0-\cot\delta\sin\delta_0\cos(\alpha-\alpha_0)}{\sin\delta_0+\cot\delta\cos\delta_0\cos(\alpha-\alpha_0)}.$$

反过来，利用下式由理想坐标也可计算赤道坐标：

$$\tan(\alpha-\alpha_0) = \frac{\xi\sec\delta_0}{1-\eta\tan\delta_0},$$
$$\tan\delta = \frac{\eta+\tan\delta_0}{1-\eta\tan\delta_0}\cos(\alpha-\alpha_0).$$

量度坐标(measured coordinate)　照相天体测量学用语.是通过坐标量度仪测量出的天体星象在底片上某一直角坐标系中的坐标 (x,y).量度底片时，调整底片在坐标量度仪上的取向，使坐标量度仪直角坐标系的 x 轴和 y 轴分别平行于赤纬圈和赤经圈，并在量度仪转像棱镜的0°和180°的两个位置量度，取其平均值作为量度坐标，以提高测量精度.量度坐标通过理想坐标 (ξ,η) 可归算为赤道坐标.

特纳法(Turner method)　亦称底片常数法.照相天体测量学中的一种方法.是通过底片常数将量度坐标 (x,y) 归算为理想坐标 (ξ,η) 的方法.19世纪由英国天文学家特纳(Turner,H.H.)提出，因而得名.由于量度坐标系和理想坐标系原点不重合，x 轴和 y 轴不正交，x 轴和 ξ 轴、y 轴和 η 轴互不平行，量度坐标受到较差大气折射和较差光行差的影响等原因，因此量度坐标与理想坐标并不一致，其关系为

$$\xi_i = ax_i + by_i + c,$$
$$\eta_i = dx_i + ey_i + f \quad (i=1,2,3),$$

式中 a,b,c,d,e,f 称为底片常数.上式称为特纳公式.特纳法的步骤为：

1. 选定底片光心的赤道坐标 (α_0,δ_0)，将定标星的 (α_i,δ_i) 换算为理想坐标 (ξ_i,η_i)(参见“理想坐标”).

2. 将定标星的理想坐标和量度坐标 (x_i,y_i) 代入特纳公式，用最小二乘法求出底片常数.

3. 将底片常数和待测天体的量度坐标 (x,y) 再代入特纳公式右端，求出待测天体的理想坐标 (ξ,η).

4. 最后，将待测天体的理想坐标 (ξ,η) 归算为赤道坐标 (α,δ)(参看“理想坐标”).

底片常数法(plate constant method)　即“特纳法”.

中天法(method of meridian observation)　一种测定时间的方法.是通过记录恒星过中天的时刻，以测定钟差的方法.中天法测时用迈耶尔公式：

$$\alpha = T + u + a\sin(\varphi-\delta)\sec\delta + b\cos(\varphi-\delta)\sec\delta + c\sec\delta$$

(对上中天观测而言)，式中 α,δ 分别为被测星的赤经、赤纬；T 为恒星中天时的钟面时；φ 为测站纬度；u 为钟差，即地方恒星时和钟面时之差；a,b,c 为仪器误差，其中 a 称方位差，为仪器水平轴偏离正东西方向之差，b 称水平差，表示水平轴的倾斜，可用水平轴上的水准器检测并改正，c 为准直差，由望远镜视准轴和水平轴不垂直引起，可用转轴观测来消除.通常观测一组星后，用最小二乘法同时解算钟差 u 和方位差 a.所用仪器有中星仪、光电中星仪和照相天顶筒等.

中星仪(transit instrument)　一种天文仪器.主要用来精确测定恒星过中天的时刻以确定钟差.仪器主要部件是望远镜、目镜测微器、挂水准器和寻星度盘，有的中星仪上还附有太尔各特水准器.望远镜垂直装在东西方向放置的水平轴上，这样望远镜

的视准轴正好在子午面内.观测时,目镜测微器和记时仪配合,记录恒星过中天的时刻,并以水平轴上的挂水准器测定水平轴与水平面的倾角,然后以迈耶尔公式求出钟差.20世纪40年代末,目镜测微器为光电记录装置所替代,以减少人差和提高观测精度.20世纪80年代末,仪器被改成自动化观测,并采用光子计数器,进一步提高了观测精度,方便了观测.利用仪器上的太尔各特水准器,中星仪还可按太尔各特法测定天文纬度.

子午环(meridian circle) 一种天文仪器.专门用于精确测定天体的赤经和赤纬.它由望远镜、目镜测微器和精密垂直度盘等组成.望远镜垂直装在水平轴上,水平轴东西方向放置,这样,望远镜只能在子午面内转动.观测天体中天时,记录下垂直度盘、目镜测微器以及记时设备的读数,就可由此精确测定天体中天时的天顶距和钟面时刻,最后计算出天体的赤经和赤纬.近代新式子午环采用光子计数器代替目镜测微器,用光学度盘或分辨率达0.05″的感应式传感器代替金属度盘,并改成全自动程序控制观测,从而减少人的影响,提高观测精度,并可观测到13等的暗星.

太尔各特法(Talcott method) 一种精确测定天文纬度的方法.方法要求:在天顶南北各选一颗恒星组成星对,其天顶距之差小于望远镜的视场,利用目镜测微器测量这对恒星的天顶距之差以求纬度.当南北两星均为上中天时,计算纬度的基本公式为

$$\varphi=\frac{1}{2}(\delta_S+\delta_N)+\frac{1}{2}(Z_S-Z_N)+\frac{1}{2}(\rho_S-\rho_N),$$

式中δ为恒星赤纬,Z为目镜测微器读数,ρ为大气折射改正,下标S,N分别表示恒星位于天顶南或北.其优点是以微小天顶距差代替天顶距的测定,以避免垂直度盘的刻度误差,并采用天顶距差的大气折射改正,以减少大气折射的误差.此法首先在18世纪由丹麦天文学家赫瑞鲍(Horrebow,C.)提出,于1857年由美国天文学家太尔各特(Talcott,A.)改进,并用于实测,故又称赫瑞鲍-太尔各特法.所用的仪器有天顶仪、中星仪或全能经纬仪等.

多星等高法(multistar equal altitude method) 同时测定天文经度(或时间)和天文纬度的方法.此法要求记录一组星过某一固定等高圈的时刻,故称多星等高法.其基本公式为

$$\cos Z=\sin\varphi\sin\delta+\cos\varphi\cos\delta\cos(T+u-\alpha),$$

式中α,δ分别为被测星的赤经和赤纬,Z为等高圈的天顶距,T为星过等高圈的钟面时,u和φ分别为钟差和纬度.理论上只要求观测3颗天顶距相等的星,就可解得三个未知量u,φ和Z.实际工作中,为提高观测精度,要求观测一组均匀分布于四个象限内通过同一等高圈的星,记录其过等高圈的时刻,用最小二乘法同时解出钟差u、纬度φ和天顶距Z.多星等高法的等高圈有60°和45°两种,前者应用较广泛.所用仪器有棱镜等高仪、光电等高仪和全能经纬仪等.

等高仪(astrolabe) 一种天文仪器.它可以同时测定时钟改正(经度)和纬度.仪器结构为一水平望远镜,物镜前放一60°棱镜,棱镜前下方放一水银盘.在星过等高圈前后,一部分星光直接射入60°棱镜的一面,再射向物镜;另一部分星光通过水银面反射后进入60°棱镜的另一面,然后射向物镜.当星到达60°高度时,从棱镜射出的两束光互相平行,在焦平面上两个像重合.记下两像重合的时刻,通过观测一组(3颗以上)位置已知的星在不同方位相继过等高圈的时刻,就可得到仪器所在地的经纬度或时钟改正和纬度.20世纪50年代,法国天文学家丹戎(Danjon,A.)在望远镜焦平面附近加一块渥拉斯登棱镜,以消除调焦带来的误差,同时增加观测记录,以提高精度.1972年和1974年,中国先后设计制造出使用光电记录器的Ⅰ型和Ⅱ型光电等高仪.20世纪80年代末,又改进为全自动的、用光子计数器记录的光电等高仪.

天顶仪(zenith telescope) 一种天文仪器.用以精确测定纬度.仪器由望远镜、目镜测微器和精密水准器等部件组成.望远镜能绕东西方向安置的水平轴旋转,因而望远镜只能在子午面内转动.观测方式是:在天顶南北各选一颗天顶距近似相等的星组成星对,望远镜先对准并观测天顶以南(北)的星,读取目镜测微器读数,然后将望远镜连同水平轴一起绕垂直轴旋转180°,再观测选定的天顶以北(南)的另一颗星,并读取目镜测微器读数.由于两星天顶距之差小于望远镜视场,所以能在同一视场中看到,由两次目镜测微器读数之差,就可测得纬度值.望远镜转轴前后在天顶距方向的微小变化,可用镜筒上的精密水准器检测并加以改正.这种仪器的优点是避免了天顶距度盘的误差和大气折射改正的误差.

天文导航(celestial navigation) 实用天文学术语.它以天体为观测目标,确定船舶、飞机和宇宙飞船的位置.早期天文导航应用航海六分仪、航空六分仪,观测太阳、月球、几颗大行星和几十颗明亮恒星,配以航海钟记录时间,由定位线图解确定经纬度.位置精度对于航海为0.1′,对于航空为1′;方位角精度航海为0.1°,航空为1°.由于受天气的限制,随后为无线电导航和人造卫星导航所取代.卫星导航是20世纪60年代以后发展起来的.卫星采用双频测速多普勒导航体制,供测站接收和定位,精度可达10m以上,且不受天气制约.宇宙飞船在行星际航行中,常以天体的亮度或特殊恒星的光谱为基准来自动导航.

六分仪(sextant) 一种观测天体的测角仪器. 1731 年,由英国天文学家哈德利(Hadley,J.)发明,19 世纪 30 年代后成为舰船上测量天体高度的主要仪器. 其特点是:不需支架固定,不需水准管置平. 操作虽简便,但精度不高,度盘成扇形,仅为整圆的六分之一,因而得名.

时间计量(time measurement) 计量时间间隔和时刻的通称. 计量时间需要选取一物体(或物质),它的运动必须是连续的、均匀的,又是可以观测到的. 时间计量包含着计量时刻和计量时间间隔两个既不相同又有联系的内容. 时刻指客观物质在其某一运动状态瞬间与时间起算点之间的时间间隔,是没有延续性的瞬间. 时间间隔指客观物质在其运动中的两个不同状态瞬间所经历的时间长度. 天文学上的时间计量专指一"日"以内的时间间隔以及选定初始历元作为时间的起算点. 人们最早是利用地球自转运动来计量时间的,其基本单位是平太阳日. 这种以地球自转为基础的时间计量系统称为世界时. 1 平太阳日的 1/86400 为世界时的 1s. 通过天文观测可测定出精确的世界时,称为天文测时. 天文观测受晴夜的限制,为了随时获得世界时,要用精密的天文时计将天文测时结果记录下来,并根据天文时计的运行规律随时指示外推的世界时,称为守时. 天文台用无线电讯号等传播手段将准确的世界时发播出去,称为授时. 20 世纪以来发现基于地球自转的世界时是不均匀的. 1958 年,国际天文学联合会决议,从 1960 年起采用历书时来代替世界时. 历书时通常由观测月球来确定,观测精度较低,无法满足科学技术的需要. 1967 年,国际计量委员会决定以原子时来代替历书时. 为了兼顾对世界时的广泛需要,国际上规定以协调世界时作为标准时间和频率发播的基础.

恒星日(sidereal day) 一种天文计时单位. 指春分点连续两次上中天经历的时间间隔. 它以地球自转周期为基础,约等于平太阳日的

$$\frac{365.2422}{366.2422},$$

即平太阳时 23h56min4.09054s. 它比相对于空间某一固定方向的地球自转周期要短 0.0084s. 春分点有平春分点和真春分点之分,恒星日有平恒星日和真恒星日之分,恒星日一般指平恒星日.

平恒星日(mean sidereal day) 天文学使用的一种计时单位. 平春分点连续两次上中天所经历的时间长度称为平恒星日. 它等于平太阳时 23h56min4.09054s. 比地球自转周期短 0.0084s. 它与平太阳日长度之比为

$$0.997269566414 - 0.586 \times 10^{-10}T,$$

与地球自转周期之比为

$$0.999999902902 - 0.59 \times 10^{-10}T,$$

T 为从 1900.0 历元起算的儒略世纪数.

真恒星日(true sidereal day) 天文学使用的一种计时单位. 真春分点连续两次上中天所经历的时间长度称为真恒星日. 由于真春分点的变化不均匀,真恒星日的长度也是变化的,与平恒星日之差最大为 0.004s.

恒星时(sidereal time) 一种天文计时系统. 指春分点的时角. 它以地球自转周期为基础,从 0～24 小时,每小时分为 60 恒星分,每恒星分再分为 60 恒星秒. 恒星时的计量单位为 1 恒星日. 恒星时从子午圈计量,不同经线上的观测者在同一瞬间所测得的恒星时是不同的,所以用地方恒星时表示. 地方恒星时 s 等于恒星赤经 α 与恒星时角 t 之和,即 $s=\alpha+t$. 观测某一恒星上中天($t=0$),则该恒星的赤经即等于该瞬间的地方恒星时. 两地的地方恒星时之差,在数值上等于两地地理经度之差. 由于岁差和章动的影响,春分点在天球上不是固定的,对应同一历元有真春分点和平春分点之别,相应地恒星时也有真恒星时和平恒星时之分.

平恒星时(mean sidereal time) 一种天文计时系统. 指平春分点的时角. 由于岁差和章动的影响,春分点在天球上是运动的. 只考虑岁差影响,以等速运动的春分点称为平春分点. 既受岁差又受章动影响的春分点称为真春分点,真春分点的时角称为真恒星时,真恒星时 s_t 和平恒星时 s_m 之差称为赤经章动,即

$$s_t - s_m = \Delta\psi\cos\varepsilon,$$

$\Delta\psi$ 为黄经章动,ε 为黄赤交角. 赤经章动变化是不均匀的,因此,真恒星时不是均匀的时间计量系统,只能用来确定时刻,不能用来计量时间间隔. 春分点在天球上的周日运行速率是地球自转角速度与春分点运行速度的合成,因此平恒星时 s_m 的变化为

$$\frac{ds_m}{dt} = \omega + m,$$

ω 为地球自转角速度,m 为赤经总岁差,

$$m = 3.07234^s + 0.00186^s T,$$

T 为从 1900.0 起算的回归世纪数,因此严格地说,平恒星时也不是均匀的时间计量系统.

真恒星时(true sidereal time) 见"平恒星时".

平太阳日(mean solar day) 一种天文计时单位. 指平太阳连续两次上中天所经历的时间间隔. 它以地球自转周期为基础,太阳圆面中心连续两次上中天经历的时间间隔则称为真太阳日. 由于太阳在天球上的周年视运动不均匀以及太阳周年视运动在黄道上运行而时角在天赤道上计量,真太阳日的长度在不断变化,最长和最短相差 51s,因此不能直接用于计量时间. 实用上,假想天球上一点均匀地沿着

天赤道作周年视运动，其运动速度等于太阳在黄道上周年视运动的平均值，该假想点称为平太阳.

平太阳(mean sun)　见“平太阳日”.

真太阳日(true solar day)　一种天文计时单位.太阳圆面中心连续两次上中天所经历的时间间隔称为真太阳日.由于太阳的周年视运动不均匀，太阳在黄道上运行而时角在天赤道上计量，使得真太阳日的长度变化不均匀，最长和最短相差51s.

平太阳时(mean solar time)　简称平时.一种天文计时系统.指以地球自转周期为基础的一种天文计时系统.用平太阳的时角来计量，以平太阳下中天的平太阳时为零时.平太阳是19世纪美国天文学家纽科姆(Newcomb, S.)引入的一个假想的参考点，它在天赤道上匀速运动，其速度与太阳在黄道上的平均运行速度相等.太阳的时角为视太阳时，视太阳时与平太阳时之差称为时差.因为平太阳是天球上一个假想的点，不能观测，实际上是直接测定恒星时再换算成平太阳时.由于平太阳时是按各地子午圈计量，因此对于某一瞬间不同地理经度地方的平太阳时是不相同的，用地方平太阳时表示.两地的地方平太阳时之差(m)在数值上等于两地的地理经度(λ)之差，即

$$m_A - m_B = \lambda_A - \lambda_B.$$

为了避免平太阳时随经度而不断变化所引起的不便，将地球划分为24个时区，每一时区经度为15°，时区中央的标准经度为0°，15°，30°等.以时区中央子午线的平太阳时为整个时区的统一时间，称为区时，又称标准时.以某一子午圈(如通过首都的子午圈)的时间作为某一区域或国家的民用时间标准，也称为标准时.区时是应用很广的标准时.中国采用东经120°的东八时区的区时为全国标准时间，又称北京时间.由于经济等原因，国家用法律规定该国(或地区)使用的时间提前一小时、半小时或几小时，称为法定时，如中国夏季曾采用提前一小时的夏令时.

法定时(legal time)　见“平太阳时”.

真太阳时(true solar time)　亦称视太阳时.一种天文计时系统.指以太阳的周日视运动为依据而建立起来的时间计量系统.真太阳指的是太阳视圆面中心.真太阳时以真太阳的时角来计量，起算点为真中午，定义为真太阳时角加12h.由于地球绕太阳运动的轨道是一椭圆，又受到月球和行星的摄动作用，反映太阳在黄道上的运动是不均匀的，又由于太阳不是在天赤道上运行，而是在黄道上运行，赤道与黄道有交角ε，即使太阳在黄道上运行是均匀的，它的赤经的增加也是不均匀的，因此真太阳时是不均匀的.为了计量时间，人们建立了均匀的与真太阳时相差不大的平太阳时系统.

视太阳时(apparent solar time)　即“真太阳时”.

时差(equation of time)　实用天文学术语.指真太阳时和平太阳时之差，即真太阳时角$t_\odot$和平太阳时角t_m之差，$\eta = t_\odot - t_m$.《中国天文年历》中的太阳表中载有每天的时差η的数值，或用下式计算：

$$\eta = \tan^2\frac{\varepsilon}{2}\sin 2\lambda - 2e\sin M + 4e\tan^2\frac{\varepsilon}{2}\sin M\cos 2\lambda - \frac{1}{2}\tan^4\frac{\varepsilon}{2}\sin 4\lambda - \frac{5}{4}e^2\sin 2M + \cdots,$$

式中ε为黄赤交角，e为地球轨道偏心率，λ为太阳平黄经，M为太阳平近点角，$M = 360° + \lambda - \omega$，$\omega$为近地点黄经.

地方时(local time)　一种计量时间的系统.以观测者所在的子午圈为基准测定的时间称为地方时.由于时角是从子午圈开始量度的，位于不同地理经线上的观测者，在同一瞬间测得的参考点(太阳、春分点)的时角是不同的，因此，每个观测者都有自己的时间，称为地方时.有地方平太阳时、地方真太阳时、地方恒星时等.在格林尼治子午线上测定的地方时称为格林尼治时间.1928年，国际天文学联合会通过决议，将由格林尼治平子夜起算的平太阳时称为世界时.为克服同一时刻各地地方时彼此不同带来的不方便，19世纪70年代后期，加拿大铁路工程师弗莱明(Fleming, W. P.)建议在全世界按统一标准划分时区，实行分区计时，即以本初子午线东西各7.5°为零时区，分别向东、向西经度每相差15°划一个时区.东西各12个时区，东十二区和西十二区相重合.各时区以中央经线的地方平太阳时为本时区的统一时间，称为区时，又称标准时.有些国家采用首都(或适中地点)的地方平太阳时为本国的统一时间，也是一种标准时.以m, T_N, M, λ, N分别表示地方平太阳时、区时、世界时、地理经度、时区顺序号，则有：

$$m = T_N + (\lambda - N) = M + \lambda,$$
$$T_N = M + N = m - (\lambda - N),$$
$$M = T_N - N = m - \lambda,$$

地理经度λ东经为正，西经为负.

区时(zone time)　见“地方时”.

标准时(standard time)　见“地方时”.

世界时(universal time)　简称UT.一种计时系统.指格林尼治地方平太阳时，也是以地球自转周期为基础的一种计时系统.1928年定名为世界时.由天文观测测定的恒星时换算得出.1956年，将直接测定的世界时记为UT0；加入因地极移动造成的地面经度变化的改正$\Delta\lambda$，则记为UT1，即

$$UT1 = UT0 + \Delta\lambda;$$

再加入因地球自转速率季节性变化的改正ΔT_s，记为UT2，即

$$UT2 = UT0 + \Delta\lambda + \Delta T_s.$$

式中 $\Delta\lambda=(x\sin\lambda-y\cos\lambda)\tan\varphi,\lambda,\varphi,(x,y)$分别为观测地点的地理经度、纬度和地极坐标；

$$\Delta T_s=0.022^s\sin 2\pi t-0.012^s\cos 2\pi t -0.006^s\sin 4\pi t+0.007^s\cos 4\pi t,$$

式中 t 为贝塞尔岁首起算的年数. 1979 年 12 月，国际上决定采用协调世界时（记为 UTC）取代世界时.

北京时间(Beijing time)　一种法定时. 中国幅员辽阔，从西到东横跨东五、东六、东七、东八、东九 5 个时区，中华人民共和国成立后，除新疆、西藏两个自治区使用东六时区的区时外，其他省市自治区都使用东八时区的区时. 1966 年以来，废除东六时区的区时，全国统一采用首都北京所在的东八时区的区时，称为北京时间.

历书时(ephemeris time)　简称 ET. 一种计时系统. 是由力学定律定义的均匀时间系统. 以历书时秒长作为时间基本单位. 历书时秒定义为 1900 年 1 月 0 日历书时 12 时瞬间回归年长度的

$$\frac{1}{31556925.9747}.$$

历书时起点以 1900 年初太阳几何平黄经为 279°41′48.04″的瞬间作为历书时 1900 年 1 月 0 日 12 时正. 将观测得到的天体位置与由历书时计算的天体历表比较，可得到相应于观测时刻的历书时. 由于难以测准太阳中心的位置，月球在天球上的视运动比太阳快 13 倍多，观测月球的精度高，实际上历书时是由观测月球位置得到的. 布朗(Brown, E.)计算并出版了改进的月球历表，把观测得到的月球位置与布朗改进的月球历表进行比较，即可得到历书时. 但由于月球视圆面较大且边缘不整齐，对其中心位置也难于测准，以致所得的历书时精度比世界时低，且需经较长的观测和归算时间才能得到，因而 1967 年第 13 届国际计量会议决议：以原子时秒长作为时间计量的基本单位. 历书时 ET 与世界时 UT2 的关系为

$$ET = UT2 + \Delta T,$$

式中 ΔT 中除包括长期变化外，还有不规则变化，只能由观测决定.

20 世纪 50 年代以来，对布朗改进的月历表又进行了三次改进计算，得到的历书时分别称为 ET0, ET1, ET2.

原子时(atomic time)　简称 AT. 一种计时系统. 是由原子内部能级跃迁所发射或吸收的电磁振荡频率建立起的时间计量系统. 是由原子钟导出的时间. 1967 年 10 月，第 13 届国际计量会议决定以原子时秒长作为国际单位制的时间单位，定义在海平面上铯原子(^{133}Cs)原子基态的两个超精细结构的能级跃迁辐射的电磁波振荡 9192631770 周所经历的时间间隔为一个原子时秒长. 原子时起点为 1958 年 1 月 1 日零时世界时. 全世界 100 多台原子钟通过各种方法比对，再由国际计量局进行归算，所求出的世界统一的原子时称为国际原子时.

国际原子时(international atomic time or temps atomique international)　简称 TAI. 一种计时系统. 即现在国际上采用的原子时系统. 秒长取海平面上的国际制(SI)秒，起点取 1958 年 1 月 1 日零时整为世界时和国际原子时的同一时刻. 根据精确计算，此时两个时间系统的时刻相差最小，仅为0.0039s. TAI 于 1972 年 1 月 1 日零时整正式启用. 不在海平面上的地点，所用时间应按广义相对论作修正.

国际制秒(international system (system international) second)　简称 SI 秒. 一种计时单位. 国际计量标准中的秒长定义为国际制(SI)秒. 由于地球自转不均匀，故用地球自转定义的时间系统平太阳时或世界时都是不均匀的. 原子钟出现后，它的均匀性和稳定性得到肯定. 在 1967 年召开的第 13 届国际计量会议上，决定了按原子时标准的秒长定义，即国际制秒或 SI 秒. 定义为：铯原子(^{133}Cs)原子基态的两个超精细结构的能级间跃迁辐射的电磁波振荡 9192631770 周所经历的时间. 上述定义与地点无关，但根据广义相对论，时间变化与铯原子钟所在地点的引力场和运动速度有关，故 SI 秒是原时单位. 另外，所取的数值 9192631770 周，是与 1900.0 年的历书时 ET 的秒长一样.

协调世界时(coordinated universal time)　简称 UTC. 一种计量时间的系统. 即以原子时秒长作为时间单位，原子时秒长累积的时刻又尽量接近世界时的一种计量时间系统. 1967 年，国际计量大会采用时间"秒"的新定义：^{133}Cs 原子基态的两个超精细结构的能级之间跃迁辐射的电磁波振荡 9192631770 周所经历的时间为 1 原子秒. 以原子秒为时间单位而建立起原子时，其起点定为世界时 1958 年 1 月 1 日零时. 由于地球自转的不均匀，近 20 年，世界时每年比原子时大约慢 1s. 为了使原子时时刻与世界时保持一致，采用增加或减少 1s（称为闰秒）的办法. 1974 年，规定原子时秒长累积的时刻与世界时 UT1 时刻之差保持在±0.9s 以内. 当原子时秒长累积的时刻与世界时 UT1 的时刻接近 0.9s 时，即加或减 1s. 闰秒日期规定为 6 月 30 日或 12 月 31 日. 1979 年 12 月，在日内瓦举行的世界无线电行政大会决定：采用协调世界时取代格林尼治时间作为无线电通信中的标准时间.

时间服务(time service)　实用天文学术语. 指为科学研究和生产提供标准时间和频率的工作. 时间服务包括测时、守时、播时、计算和出版时号改正数等. 主要分为世界时服务和原子时服务. 世界时服

务指从事时间服务的天文台利用大量的天文测时资料进行误差处理，求得精确的世界时，每天先按世界时的近似外推值用无线电时号的形式发播出去，再根据其后测算的精确世界时对过去已发播的近似值进行修正. 时号改正数是世界时服务的最后成果. 将许多天文台所定的时号改正数进行综合处理，或者直接利用这些天文台的天文测时资料进行综合处理，这样得到的时号改正数称为综合时号改正数，作为世界时标准. 在采用原子时以后，无线电时号一般均按协调世界时或原子时发播. 原子时服务指以原子钟为基础，用无线电时号将协调世界时或原子时发播出去. 通过各种时间比对的手段，将各天文台的原子钟所显示的原子时进行比较，经过综合分析处理得到协调世界时或原子时的标准. 国际时间局所提供的原子时标准称为国际原子时. 国际时间局定期发表国际协调世界时与各天文台提供的协调世界时的差. 中国从 1959 年起建立综合时号改正数系统. 现在由上海天文台和陕西天文台分别根据原子钟发播 BPV 和 BPM 等时号.

历法(calendar)　一种计量长时间间隔的计时系统. 是规定年、月、日的长度和它们之间的关系，制定时间序列的法则. 也可以理解为年、月、日三种记时单位排列、组合的方式. 在以往的历法中，日的长度是按太阳的视运动定出的，年的月数和日数，月的日数是按季节和天象确定的，这样，年和月的日数就不一定是一个有理数，如一回归年为 365.24220…日，一朔望月为 29.53059…日，为了使人们生活方便，历法必须使年和月的日数为整数，因此历法本身有一定的人为因素.

目前各国使用的历法有阳历(如公历)、阴阳合历(如中国传统历法夏历)、阴历(回历)等.

阳历(solar calendar)　亦称公历. 历法的一种. 是以太阳周年视运动为依据制定的，即历年的平均长度约等于回归年；历月的长度由人为规定，与朔望月无关. 古罗马儒略·恺撒(Caesar, G. J.)于公元前 46 年改历规定：历年的长度为 365 日，每 4 年增加 1 日，为 366 日(闰年)，即历年的平均长度为 365.25 日，故称儒略历，又称旧历. 1582 年，罗马教皇格里哥里十三世(Gregorius, Ⅷ)修改儒略历的置闰规则，规定对世纪年，只有被 400 除尽的才算作闰年，使历年的平均长度为 365.2425 日，故称格里历，又称新历. 目前，通用于世界绝大多数国家的历法即为格里历.

公历(gregorian calendar)　即“阳历”.

儒略历(Julian calendar)　见“阳历”.

格里历(Gregorian calendar)　见“阳历”.

阴历(lunar calendar)　历法的一种. 是根据月相变化周期(朔望月)制定的一种历法. 历月的平均长度大约与朔望月的长度相等. 规定单数的历月为 30 日，双数的历月为 29 日，每月平均 29.5 日，并以新月始见为月首. 12 个月为一年，共 354 日. 由于 12 个朔望月的长度是 354.36708 日，比阴历历年长 0.36708 日，30 年共长 11.0124 日. 阴历每 30 年中加 11 个闰日，规定在第 2，5，7，10，13，16，18，21，24，26，29 各年的 12 月底加一个闰日，有闰日的年称为闰年，计 355 日. 伊斯兰教国家使用的回历即是一种阴历. 民间常把中国传统历法夏历称为阴历，而把通用的公历称为阳历.

阴阳历(luni-solar calendar)　历法的一种. 兼备阳历和阴历的特点. 即历年以季节变化周期回归年为依据，历月按月相盈亏变化周期朔望月来制定. 即历月的平均长度约等于朔望月，历年的平均长度约等于回归年. 大月 30 日，小月 29 日，一般大小月相间. 平年 12 个历月，每 19 年加 7 个月，增加的月称闰月，有闰月的年称为闰年，即闰年为 13 个月. 中国传统历法除天历外，均为阴阳历. 中国 1912 年后改用公历，但未完全废除传统使用的阴阳历，后者被称为夏历或农历，但常被民间称为阴历. 在阴阳历中引入 24 节气，每个月都有固定的中气，19 年中有 7 个月没有中气，规定该月作为这一年的闰月，用上月的名称加一个“闰”字.

回归年(tropical year)　天文学术语. 指四季更迭的周期. 即日面中心连续两次通过春分点所经历的时间间隔. 回归年的长度等于

$$365.24219879 - 0.00000614(t - 1900)$$

平太阳日，t 是年份数字. 历法中的“年”即按此来安排的. 它比地球围绕太阳平均公转周期——恒星年约短 20 分 23 秒，后者等于 365.25636 平太阳日，即 365 日 6 时 9 分 9.5 秒. 地球连续两次经过地球和太阳之间最近点的平均时间间隔称近点年，等于 365.25964 平太阳日. 太阳在天球上连续两次通过月球轨道的升交点所经历的时间间隔称交点年，又称食年，等于 346.62003 平太阳日，即 346 日 14 时 52 分 53.1 秒.

恒星年(sidereal year)　天文学术语. 指地球围绕太阳运行的平均公转周期. 即太阳在黄道上连续两次经过某一固定点(或某一固定春分点)的时间间隔. 它与回归年仅差一黄经岁差. 恒星年长度

$$t' = 365.25636042 + 0.00000011\,T$$

平太阳日，T 为从 1900 年 1 月 0 日起算的儒略世纪数，每儒略世纪等于 36525 平太阳日. 恒星年的长度是变化的，它将越来越长.

交点年(nodical year)　见“回归年”.

近点年(anomalistic year)　见“回归年”.

恒星月(sidereal month)　天文学术语. 指月球围绕地球平均运转的周期. 即月球相对于空间某一

固定方向的平均绕地球公转周期，即月球绕地球运行连续两次通过同一点所经历的时间. 平均长度为27.32166 日. 月球连续两次经过月球和地球之间最近点的平均时间间隔称为近点月，长度为 27.555日. 月球相继两次经过黄道和白道交点的平均时间间隔称为交点月，长度为 27.21222 日. 月球平黄经增加 360°所需要的时间称为分至月，它的长度为27.322 日.

朔望月(synodic month)　天文学术语. 指月相盈亏变化的周期. 即月球相对于太阳的平均绕地球公转的周期. 月球从朔(望)到下一次朔(望)的平均时间间隔为29.53059 日. 中国传统历法中的历月就是以朔望月为基础而编算的. 朔望月的长度 S 可由会合运动方程式

$$\frac{1}{S}=\frac{1}{T}-\frac{1}{E}$$

求出. 式中 T 为恒星月，月球绕地球公转从恒星间某一点再回到同一点所需的时间为 27.32166 日；E 为恒星年，太阳周年视运动从恒星间一点回到同一点所需要的时间为365.25636 日.

交点月(nodical month or draconitic month)　见“恒星月”.

近点月(anomalistic month)　见“恒星月”.

分至月(tropical month)　见“恒星月”.

儒略日(Julian day)　简称 JD. 天文学术语. 指天文学上用日来计算长时间间隔的一种计时系统. 取公元前 4713 年 1 月 1 日格林尼治平太阳时 12 时为起点，顺序连续计日. 如 1990 年 1 月 1 日零时的儒略日为 2447892.5，1990 年 10 月 1 日零时的儒略日为 2448165.5. 这一记法是 16 世纪荷兰的一位教授斯卡里格尔(Scaliger)所提出的，为纪念他的父亲儒略(Julius)而命名为儒略日. 由于儒略日数字位数太多，1973 年，国际天文学联合会决定采用简化儒略日，记为 MJD，

$$\mathrm{MJD}=\mathrm{JD}-2400000.5.$$

MJD 的相应起点是 1858 年 11 月 17 日世界时零时.

天文地球动力学(astrogeodynamics)　天体测量学和地学交叉的一门学科. 它运用天文观测手段，研究地球的各种运动及产生的机制. 地球的运动包括地球自转和公转，还有地球内部、地壳、水圈和大气圈的物质运动. 研究产生这些运动的机制涉及地球的内部结构、物理性质、物质运动、重力场结构及其变化、地球水圈和大气圈的大规模物质运动、空间引力场和地磁场作用、地球和太阳系的起源与变化. 随着新的天文观测技术和空间技术的发展，测定地球自转和地壳运动的精度已达到厘米级，因而推动了天体测量学和地学的密切结合，促进了天文地球动力学的发展. 它的主要研究课题有：地球自转速度变化和极移的规律与机制、板块运动及其机制、固体潮、地球弹性参数和地球重力场的测定、地球内部结构对地球运动的影响等.

地球自转(rotation of the earth)　地球的一种周期性旋转运动. 地球本体绕自转轴自西向东的周期性旋转称为地球自转. 自转平均角速度为7.292×10^{-5}rad/s. 自转的周期——日，自古以来人们用作计量时间的基本单位. 20 世纪以来，随着测时、守时精度的提高，发现地球自转速度不均匀，后来人们就先后以历书时和原子时为计量时间的基本单位. 地球自转速度的变化有三种：

1. 长期减慢. 在一个世纪内使日长增长1～2ms. 主要原因是日、月引力引起的潮汐摩擦造成地球自转角动量减少；其次是地幔与地核间角动量的交换、海平面和冰川的变化等.

2. 不规则变化. 例如，在几十年到更长时间中，有每年小于$\pm5\times10^{-10}$的相对变化；几年到十几年时间中，有每年小于$\pm8\times10^{-9}$的相对变化；几周到几个月中，有每年小于$\pm5\times10^{-8}$的相对变化. 其原因可能同地幔与地核间角动量交换，或海平面变化和风的作用等有关.

3. 季节性周期变化. 包括振幅约 20～25ms 的春慢秋快的周年变化和振幅约为 9ms 的半年变化. 其原因与风的季节性变化及太阳潮汐有关.

由于天文测时精度的提高，在地球自转速度变化中，还发现一些微小短周期变化，振幅约 1ms，其原因可能与月球潮汐有关. 地球自转除速度变化外，还有方向变化. 自转轴在空间的变化，就是岁差和章动. 自转轴在地球本体内的运动，就是极移. 关于地球自转的研究，对天文学、地学和古生物学都有重要影响，现在它已发展成为一门交叉性的新学科.

极移(polar motion)　天文地球动力学术语. 地球瞬时自转轴在地球本体内的运动造成地极沿地面的移动，称为极移. 1765 年，瑞士数学家欧拉(Euler，L.)在地球为刚体的假定下，从理论上预言极移的存在. 1888 年，德国天文学家屈斯特纳(Küstner，K. F.)才从实测纬度变化中发现极移. 1891 年，美国天文学家张德勒(Chandler，S. C.)进一步指出：地极移动中主要有两种周期，一个是周期约为 14 个月、振幅为 0.1″～0.3″的张德勒周期，这是非刚体地球的自由摆动；另一个是周期为 1a、振幅约 0.1″的周年周期，主要是大气引起的受迫摆动. 此外，还有长期极移和各种短周期的复杂运动. 极移的形成有两方面因素：其外部因素是日月引力、大气和海洋的作用；而内部因素则与地球内部结构有关. 极移使地球表面各点的经度、纬度和方位角发生变化，实测工作中，需根据极移数值加以修正. 尽管地极移动的轨

迹异常复杂，但其移动幅度不大，地极在地面上的移动范围不超过$(30\times30)m^2$. 历史上曾用观测地面点的纬度变化来研究极移，现代则使用人造卫星多普勒观测以及激光技术来研究极移.

极移服务(polar motion service) 天文地球动力学术语. 指组织一些天文台进行协作，用天文学方法测定地极移动，向天文、大地测量、空间科学等方面系统地提供地极坐标的服务工作. 国际上进行极移服务的系统先后有以下几个：

1. 国际纬度服务(I LS). 成立于 1899 年，由北纬 39°08′的几个纬度台站组成. 用相同的仪器和观测方法，观测相同的恒星，以最大限度地减小各台站的公共非极移变化. 所得地极坐标的精度在 1m 以上.

2. 国际极移服务(IPMS). 1962 年成立. 将全世界各天文台站提供的测时、测纬结果进行综合，提供系统的地极坐标. 所得地极坐标的精度约在 1m 左右.

3. 国际时间局(BIH)地极坐标系统. 1955 年开始工作. 当时因国际纬度服务提供地极坐标的速度太慢，所以成立国际时间局的快速纬度服务，提供地极坐标，以满足时间服务的需要. 1968 年，国际时间局采用国际习用原点 CIO 为地极坐标原点，并用测时、测纬结果联合计算地极坐标. 1972 年起，在系统中加入人造卫星多普勒方法观测的资料，所得地极坐标精度好于 1m.

4. 达尔格伦地极监视服务(DPMS). 1967 年成立. 根据十多个观测站的多普勒观测结果，相对国际习用原点确定地极坐标. 所得精度高，提供结果的时间短. 但它和经典技术所定结果之间存在系统差，且长期稳定性也有待研究.

5. 苏联标准时刻系统(ЭВ)的极移服务. 1953 年成立. 用奥尔洛夫单站纬度观测计算地极坐标的方法，综合苏联五个台的纬度观测，给出地极坐标.

6. 中国极移服务. 1964 年起，由天津纬度站提供相对历元平极的地极坐标. 1971 年后，继续综合中国各天文台的测纬观测，提供相对历元平极的地极坐标. 1977 年后，应用国内外纬度观测资料，取 1968.0 的瞬时极为原点，计算得到的地极坐标，称为极原点系统(JYD). 地极坐标的精度约 1m.

7. 国际地球自转服务(IERS). 由国际大地测量学和地球物理学联合会(IUGG)以及国际天文学联合会(IAU)联合成立，并于 1988 年起正式代替原来的国际极移服务和国际时间局关于地球自转的工作. 根据由甚长基线干涉测量、激光测月和激光测卫的测时和测纬资料，综合处理，给出相对国际习用原点的地极坐标. 所得精度在 10～20cm 以内，是目前国际上最广泛采用的地极坐标系统.

人造卫星多普勒观测(satellite Doppler measurement) 人造卫星定轨、地面测站定位和精测地极坐标的一种新方法. 当卫星飞越测站上空，与测站距离不断变化，致使接收到由卫星发射的高稳定信号频率也在不断变化，称为多普勒频移. 一般采用人造卫星发射两个相干频率，来消除电离层影响. 接收多普勒频移，实用价值很高，具有全天候、遍全球、速度快、精度高、仪器便携等优点. 它用来定轨和定位的精度，可达 10m 以内，甚至 1m. 所以又可用于精测地球引力场模型并广泛用于导航、勘察和大地测量等领域. 美国自 20 世纪 60 年代起开发子午仪卫星，用于海洋导航. 日本也在研制相似的导航卫星.

人造卫星激光测距(satellite laser ranging) 天文地球动力学的一种测定距离的方法. 是以激光为光源，精确测定地面测站到人造卫星(装有后向反射器)距离的一种新方法. 特点是不需要测角，从而避免大气折射的影响. 用望远镜接收返回的激光信号，并准确记录到达卫星的往返时间，记时精度达 0.1ns，测距精度达 10cm 量级，甚至更高. 再根据卫星轨道要素，解算测站地心坐标或测定数千公里长的地面基线. 自 1963 年开始应用，后被推广，美、法等国先后发射 10 多颗此类专用卫星，曾有 20 多个测站进行国际合作研究. 除大地测量学任务外，还广泛应用于地球物理学、天体测量学以及地震预报等领域的研究.

射电天体测量学(radio astrometry) 天体测量学的一个新分支. 它利用射电干涉技术测量天体的位置. 20 世纪 60 年代后期，由于甚长基线干涉仪的试验成功，使分辨率和测量精度大为提高，从而使射电天体测量迅速发展成为一门独立的学科. 射电天体测量方法有两种，即甚长基线干涉测量和连线干涉测量. 现在，射电天体测量的精度已超过了光学天体测量的精度. 射电天体测量有以下优点：不论角度大小，测量精度基本相同；大气折射影响小；可自动以地球瞬时自转轴为参考，绝对测定赤纬；可根据观测值同时求出天体位置、基线参数和仪器参数，减少后两者对天体位置的影响；可在白天和云雾天气下观测. 利用射电天体测量，可提供惯性参考系，精确测定某些天文常数，测定世界时、极移和地面点坐标，以及研究天文地球动力学的有关问题.

空间天体测量学(space astrometry) 天体测量学的一个新分支. 它利用现代空间技术将望远镜送到地球外层空间以测量天体的位置. 地面观测受到各种外界因素的干扰，其中主要是大气和重力的影响. 和地面观测相比，空间天体测量有以下优点：由于没有大气，可消除大气折射带来的误差；由于不存在大气吸收，能在可见光和其他波段进行天体测量；由于没有大气漫射，天空背景黑暗，仪器可连续

进行观测;由于没有大气闪烁,成像质量好,暗星和角距离很小的双星都可观测;由于没有重力,望远镜镜筒和仪轴不会弯曲.空间天体测量只能测定天体间的相对位置,必须通过射电观测等方式才能得到绝对位置.空间天体测量可用于改进基本星表系统,并把它与由射电天体测量建立的惯性参考系联系;测定恒星自行和视差,进而精确测定恒星的绝对星等和银河系的距离尺度;测定暗星和星系核的角直径;测定近距分光双星,确定天体质量.

天体力学

天体力学(celestial mechanics)　天文学的一个古典分支.它研究天体的运动和形状.其中运动指力学运动,包括自转和在空间中的移动.而自转与形状又有密切关系.人们所研究的天体范围在逐步扩大,现在已包括太阳系中全部自然天体和人造天体,也包括各种恒星系统,但仅限于力学运动研究.有的国家把这个学科称为动力天文学.长期以来,研究天体的力学运动都是以牛顿力学为基础,现代观测技术使天体定位观测精度迅速提高后,按牛顿力学建立起来的天体运动理论已不能准确解释观测现象,以爱因斯坦(Einstein, A.)的广义相对论为基础的天体运动理论,正在建立中.

古代的天文学以研究天体视运动及其应用为主,对行星和月球的长期观测进行研究,逐步建立了太阳系的运动学模型;到17世纪初出现的开普勒三大定律,使这种运动学模型达到完美程度.但是行星为什么按开普勒三大定律运动?虽然只有一些猜测,却表明正孕育着新学科的出现.

1.诞生和奠基.牛顿(Newton,I.)把17世纪出现的动力学萌芽和天体运动的开普勒三大定律结合起来,建立了牛顿力学,包括万有引力定律和牛顿三大运动定律.用于建立行星和月球的牛顿动力学理论,不仅解释了行星按开普勒三大定律运动的原因是太阳和行星间存在万有引力,还指出此定律仅为近似结果.牛顿在1687年正式出版的《自然哲学的数学原理》中宣告天体力学的诞生.人们从研究天体视运动发展到研究天体的真运动,在天文学发展史中是一个重大的飞跃.但是牛顿在此时尚未发明微积分,主要结果是用初等数学和几何方法描述的,因而无法建立此学科的系统理论和方法,甚至天体力学这个学科名词都未正式提出.天体力学的奠基过程是与数学分析和分析力学的建立与发展一起进行的.主要奠基者都是创立数学分析和分析力学的开拓者,他们的代表有欧拉(Euler,L.)、克莱罗(Clairaut,A.-C.)、达朗贝尔(d'Alembert,J. le R.)、拉格朗日(Lagrange,J.-L.)和拉普拉斯(Laplace,P.-S.)五位.最后由拉普拉斯集其大成,系统地建立起经典天体力学的主要内容——摄动理论.他在1798年出版的一本著作中,正式提出了天体力学的学科名称,他指出:"用万有引力定律研究太阳系及宇宙里类似系统中的固体和流体的平衡形状和运动理论,组成了天体力学."他在1799~1825年间出版的五卷16册历史性名著《天体力学》就是经典天体力学的代表作.至此天体力学正式奠基.不久,1846年由天体力学理论预言而发现海王星,标志着天体力学日趋成熟,成为当时天文学的主流.

2.19世纪后期的新发展.到19世纪后半叶,由于航海和大地测量的发展,要求更准的天文历表;物理方法,特别是照相和分光方法用于天文观测,可观测到亮度更暗的天体,并可测得视向速度;小行星和彗星大量发现,传统理论不适用,要求更好的理论.这些促使天体力学的发展形成一个新高潮,一直延续到20世纪初.其中主要代表人物是:德洛内(Delaunay,C.E.)、汉森(Hansen,P.A.)、希尔(Hill,G.W.)、纽科姆(Newcomb,S.)和庞加莱(Poincaré,(J.-)H.).前三位是建立新摄动理论的主要代表,他们的理论至今仍在应用和发展;纽科姆是历书天文学的创立者;庞加莱是这一时期的集大成者,他不仅把摄动理论(分析方法)进行总结提高,还创立了天体力学定性理论,并发展了天体形状和自转理论.他的历史性名著《天体力学新方法》(Méthodes Nouvelles de la Mécanique Céleste)三卷(1892~1899年出版),是这段历史时期的主要代表作.

此外,科威耳(Cowell,P.H.)等人为计算彗星位置而建立的数值方法,逐渐发展成为天体力学数值方法.

3.现代天体力学的主要领域.20世纪50年代以后,由于人造天体出现,成为新的研究对象;计算技术迅速发展,促进了天体力学的各个领域的发展;几种新观测技术出现(原子频率、激光测距、多普勒测速、甚长基线干涉测量(VLBI)、空间测量等),使天体观测的定时、定位精度大幅度提高,要求改善天体的运动理论,这些促使天体力学的发展形成又一个新的高潮.在这次高潮中,不仅充实发展了原有研究领域,还建立了新领域.从近期的进展情况看,现代天体力学的主要研究领域有:

1)摄动理论.用分析方法将天体的受摄运动方程解出为级数形式,为经典天体力学的主要内容.现代已用计算机推导级数解,并与数值方法结合而形成半分析方法.

2)天体力学定性理论,又称定性方法.是直接从天体运动方程讨论出轨道性质和时间趋于无穷时的运动特性,其发展很快.

3)天体力学数值方法.用数值方法解天体运动

方程,并已广泛用于其他领域.

4）天体的形状和自转理论.用流体或固体和弹性体作为天体模型,讨论它们在自转时的平衡形状及其稳定性.

5）天文动力学.讨论各种类型人造天体在发射段、自由飞行段及返航段中的动力学课题,是人造天体出现后建立的新领域.

6）历书天文学.研究将天体运动理论具体应用到编制历表的方法以及有关的天文参考系和天文常数系统.

7）多体问题.实质上是质点组动力学,只是以太阳系或恒星系统的天体为背景,讨论它们在引力作用下的运动规律.

8）恒星系统动力学.讨论双星、聚星、星团、星系甚至宇宙的动力学课题,作为天体力学的领域.它不用统计方法.

9）动力演化.讨论天体或天体系统在动力学作用为主阶段的演化规律.

10）相对论天体力学.在广义相对论或其他新引力理论基础上建立天体的运动理论.

动力天文学(dynamic astronomy)　即“天体力学”.

恒星系统动力学(stellar system dynamics)　见“天体力学”.

二体问题(two-body problem)　两个质点在引力作用下的运动问题.是在忽略其他质点的作用,两质点间的作用仅为引力作用的假定下,已知两质点某时的相对位置、质量、运动速度,求解任一时刻它们的位置和运动速度的问题.二体问题是太阳系天体运动理论的第一次近似结果.在二体问题中,太阳系天体的运动轨道为一种圆锥曲线,天体运动遵从开普勒行星运动定律(第三定律略有改变).在以太阳为原点的直角坐标系中,行星的运动方程为

$$\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}=-\frac{\mu}{r^2}\frac{\boldsymbol{r}}{r},$$

式中 r 为行星的向径,$\mu=G(M+m)$,G 为万有引力常数,M 为太阳质量,m 为行星质量.其轨道积分为

$$r=\frac{p}{1+e\cos f},$$

为一个以原点为焦点的圆锥曲线方程.$p=a(1-e^2)$ 为半通径,a 为轨道半长径,e 为轨道偏心率,f 为真近点角,极轴取近日点方向.

万有引力定律(law of gravitation)　物体间相互作用的一条定律.任何物体之间都有相互吸引力,这个力的大小与各个物体的质量成正比,而与它们之间的距离的平方成反比.以 m,M 分别表示两物体的质量,以 r 表示其间的距离,则两物体之间的引力 F 的大小可表示为

$$F=G\frac{mM}{r^2},$$

式中 G 为比例常数,称为万有引力常数,G 的值随所选用的计算单位制的不同而不同.若用米·千克·秒制,则

$$G=6.672\times10^{-11}\mathrm{m}^3\cdot\mathrm{kg}^{-1}\cdot\mathrm{s}^{-2}.$$

在天文学中常用天文单位作长度单位,太阳质量为质量单位,平太阳日为时间单位,此时 $G=k^2$,k 称为高斯引力常数,$k=0.01720209895$.在理论研究中常采用 58.132441 平太阳日为时间单位,则 $G=1$.万有引力定律是英国科学家牛顿(Newton,I.)在 1687 年出版的《自然哲学的数学原理》一书中首先提出的,是根据力学三大基本定律和开普勒行星运动三定律分析、推演而归纳出来的.从万有引力定律可推导出开普勒定律,并修正开普勒第三定律 $a^3/T^2=$ 常数为 $a^3/T^2(M+m)=$ 常数,M 和 m 分别表示太阳和行星质量,T 和 a 分别表示行星绕日公转周期和轨道半长径.万有引力定律是天体力学的理论基础.对于阐明月球和行星运动取得了辉煌的成功.1859 年以后发现了水星近日点剩余进动、星光在太阳附近弯曲等万有引力定律无法解释的现象,20 世纪以来,爱因斯坦(Einstein,A.)等曾提出不同的新的引力理论,如广义相对论等,能更准确解释这些现象.

开普勒(行星运动)定律(Kepler law)　行星运动的三条定律.是德国天文学家开普勒(Kepler,J.)分析天文观测资料后得到的.第一定律是太阳系中各行星均沿太阳为其一焦点的椭圆轨道运动,即行星轨道方程可写为

$$r=\frac{p}{1+e\cos f},$$

p,e 分别为行星轨道的半通径、偏心率,轨道半长径为 a,$p=a(1-e^2)$,f 为真近点角;第二定律是自太阳到行星的半径向量在相等的时间内扫过相等的面积,即

$$\frac{\mathrm{d}A}{\mathrm{d}t}=\frac{r^2\theta}{2}=\frac{h}{2},$$

A 为面积,r 为向径,θ 为幅角,h 为常数,

$$h=2\frac{\pi ab}{T}=\frac{2\pi a^2\sqrt{1-e^2}}{T},$$

T 为轨道周期;第三定律是各行星运动的周期的平方与其轨道半长径的立方成正比,即

$$\frac{a^3}{T^2(M+m)}=\frac{G}{4\pi^2},$$

G 为万有引力常数,a 表示行星轨道半长径,T 表示行星轨道周期,M,m 分别为太阳、行星的质量.

太阳系内的引力定律(gravitational law in solar system)　太阳系内各天体之间作用力遵循的规

律.牛顿(Newton,I.)首先证明在与距离平方成反比的引力作用下,行星运动符合开普勒定律.后来亚当斯(Adams,J.C.)和勒威耶(Le Verrier,U.J.J.)以1°的精度预报海王星的位置,这是牛顿万有引力定律的辉煌成果.随着观测技术的发展,以万有引力定律为基础的大行星运动理论的观测值与理论值之间出现系统偏差,引起人们怀疑平方反比定律

$$\frac{Gmm'}{r^n}$$

中的 n 是否正好等于2?贝特朗(Bertrand,G.)认为应大于2,纽科姆(Newcomb,S.)将 $n=2+1.574\times10^{-7}$ 用于天文年历中.黎曼(Riemann,(G.F.)B.)提出万有引力为

$$F=\frac{Gmm'}{r^n}\left\{1+\frac{1}{c^2}\left[\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2+\left(\frac{dz}{dt}\right)^2\right]\right\}.$$

另一个引人注意的问题是万有引力常数 G 是否为常数.狄喇克(Dirac,P.A.M.)认为 G 与宇宙年龄成反比.现已证实爱因斯坦(Einstein,A.)的广义相对论能更好地解释天体运动现象.

引力常数(gravitational constant) 天体力学中的重要常数.指牛顿万有引力定律中所定义的常数,记为 G,又称万有引力常数.用地面实验能测准它,至今仍取1973年的值,

$$G=6.672\times10^{-11}\mathrm{m^3\cdot kg^{-1}\cdot s^{-2}},$$

只有四位有效数字.但天文方法可用精确的观测资料定出 G 和某天体质量的乘积,如 GE,GS,其中 E,S 分别为地球和太阳的质量.GE 又称为地心引力常数,GS 称日心引力常数.至今仍用国际天文学联合会1976年决定的数值:

$$GE=3.986005\times10^{14}\mathrm{m^3/s^2},$$

$$GS=1.32712438\times10^{20}\mathrm{m^3/s^2}.$$

高斯常数(Gauss constant) 以天文学单位系统表示的万有引力常数.若用下面天文学单位系统:以平太阳日为时间单位,太阳质量为质量单位,日地平均距离为长度单位;再把万有引力常数 G 的平方根记为 k,则称 k 为高斯常数,因为德国数学家高斯(Gauss,C.F.)于19世纪初首先使用而得名.在讨论太阳系行星运动时,用天文学单位更方便,特别是再改用 $1/k$ 平太阳日作时间单位后,相应万有引力常数值为1.纽科姆(Newcomb,S.)于1896年准确定出 $k=0.01720209895$.国际天文学联合会1976年决定的天文常数系统中,高斯常数作为定义常数.

二体问题的解(solution of two-body problem) 两质点在引力作用下运动问题的解.用常微分方程组表示,在惯性坐标系中为12阶,相对运动为6阶.其解要通过积分来求出,两种情况都是可积系统.以相对运动为例,通解要包含六个相互独立的积分常数,用它们可把微分方程组的基本变量(如坐标和速度)表示为时间和积分常数的函数,一般形式为:

$$x=\frac{p[\cos\Omega\cos(f+w)-\sin\Omega\sin(f+w)\cos i]}{1+e\cos f},$$

$$y=\frac{p[\sin\Omega\cos(f+w)+\sin\Omega\sin(f+w)\cos i]}{1+e\cos f},$$

$$z=\frac{p\sin(f+w)\sin i}{1+e\cos f},$$

其中(x,y,z)为二体相对位置的直角坐标,式中只有 f 为时间的函数,具体关系与轨道曲线有关,椭圆轨道情况为

$$\cos f=(\cos E-e)/(1-e\cos E),$$

$$\sin f=\sqrt{1-e^2}\sin E/(1-e\cos E),$$

$$E-e\sin E=n(t-T)=nt+M_0.$$

积分常数就是轨道要素,即 a,e,i,Ω,ω,T(或 M_0).式中出现的 p,n 不是独立的常数,分别为

$$p=a(1-e^2),\quad n^2a^3=G(M+m),$$

其中 G 是万有引力常数,M,m 分别为二体质量,f,E 称为真近点角和偏近点角.

面积积分(area integral) 动量矩守恒定律的表现形式之一.是二体问题微分方程的一种积分.在直角坐标系中,二体相对运动的面积积分为

$$\frac{y\mathrm{d}z}{\mathrm{d}t}-\frac{z\mathrm{d}y}{\mathrm{d}t}=h\sin\Omega\sin i,$$

$$\frac{z\mathrm{d}x}{\mathrm{d}t}-\frac{x\mathrm{d}z}{\mathrm{d}t}=-h\cos\Omega\sin i,$$

$$\frac{x\mathrm{d}y}{\mathrm{d}t}-\frac{y\mathrm{d}x}{\mathrm{d}t}=h\cos i,$$

其中 Ω,i 为轨道要素,h 称为面积常数,h 与轨道要素 a,e 有下面简单关系:

$$h^2=G(M+m)a(1-e^2).$$

在轨道平面上,用极坐标(r,f)表示的面积积分为

$$r\frac{\mathrm{d}f}{\mathrm{d}t}=h,$$

式中 h 就是向径 r 在单位时间内扫过的面积的两倍,f 为真近点角.

轨道积分(orbital integral) 二体相对运动方程的一个积分.如用极坐标(r,f)表示相对位置,其中极轴取近点方向,f 就是真近点角,则轨道积分为圆锥曲线方程

$$r=\frac{a(1-e^2)}{1+e\cos f}=\frac{p}{1+e\cos f},$$

式中 a,e,p 分别为圆锥曲线的半主径、偏心率和半通径.

活力公式(vis viva formula) 天体力学中求解二体问题所得到的一个积分.它表示总能量守恒关系式,即

$$V^2=\mu\left(\frac{2}{r}-\frac{1}{a}\right),$$

V 和 r 分别为一天体相对于另一天体的速度和距离,$\mu=G(M+m)$,G 为万有引力常数,$M+m$ 为两

天体质量之和，a 为常数，表示椭圆轨道的半长径. 此式表示质量为 m 的运动天体的动能 $\left(\frac{1}{2}mV^2\right)$ 和引力势能 $\left(-\frac{\mu m}{r}\right)$ 之和为一常数 $\left(-\frac{\mu m}{2a}\right)$. 又称活力积分."活力"一词来源于拉丁文 vis viva，其物理意义为"动能".

轨道要素(orbital elements)　亦称轨道根数. 描述天体或人造卫星运行轨道的各种参数. 它们表示轨道的形状、大小、轨道平面在空间的位置和天体某一时刻在轨道上的位置. 行星和人造卫星的轨道要素为 6 个，轨道为抛物线的彗星轨道要素为 5 个，双星的轨道要素为 7 个. 椭圆轨道的要素为：半长径 a，偏心率 e，轨道倾角 i，升交点黄经 Ω，近日(地)点角距 ω 和过近日(地)点时刻 τ；抛物线轨道要素为：过近日点时刻 T，升交点黄经 Ω，近日点角距 ω，轨道倾角 i 和近日距 q；在双星系统中，伴星相对于主星的运动轨道一般为椭圆，其轨道要素为：半长径 a，偏心率 e，周期 p，交点位置角 Ω，近星点对于交点的经度 ω，观测所提供的视椭圆平面与真椭圆面的交角 i 和过近星点时刻 T.

轨道根数(elements of orbit)　即"轨道要素".

拉普拉斯积分(Laplace integral)　二体相对运动方程的一组积分. 拉普拉斯(Laplace，P.-S.)首先提出，故得此名. 如用向量表示，则有简单形式：

$$\frac{\boldsymbol{r}\times \mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\times\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}+\frac{G(M+m)\boldsymbol{r}}{r}=\boldsymbol{F},$$

其中 $\boldsymbol{r}(x,y,z)$ 为二体相对位置向量，M,m 分别为二体质量，G 是万有引力常数. 向量 $\boldsymbol{F}$ 称为拉普拉斯向量，它的三个分量是积分常数.

平近点角(mean anomaly)　描述天体沿椭圆轨道运动的一种参数. 为椭圆运动中的一种位置角，常用 M 表示. 由近点起算，随时间均匀变化，其变率为椭圆运动的平均角速度 n，故为时间的线性函数，关系为

$$M=nt+M_0=n(t-t_0),$$

其中 M_0 为时间 $t=t_0$ 时的平近点角，是轨道要素之一，t_0 为天体经过近点时刻，也是一个轨道要素. 在椭圆运动中，常用平近点角取代时间 t 作自变量.

偏近点角(eccentric anomaly)　描述天体沿椭圆轨道运动的一种参数. 即椭圆运动的位置角之一，常用 E 表示. 它也是由近点起算，同平近点角 M 的关系为

$$E-e\sin E=M=nt+M_0,$$

此式又称为开普勒方程，式中 e 是椭圆轨道的偏心率. 用偏近点角 E 作位置角后，椭圆运动的向径 r 同它有简单关系：

$$r=a(1-e\cos E),$$

式中 a 是椭圆轨道的半长径，E 由开普勒方程同时间发生联系，故有时也用偏近点角 E 作自变量.

真近点角(true anomaly)　描述天体在圆锥曲线轨道上运动的一种参数. 为圆锥曲线轨道的位置角之一，常用 f 表示. 由近点起算，表示沿运动方向到任一时刻位置所转的角度. 它同向径 r 的关系就是轨道积分，即

$$r=\frac{a(1-e^2)}{1+e\cos f}.$$

在椭圆轨道情况，它同偏近点角 E 的关系为

$$r\cos f=a(\cos E-e),$$

$$r\sin f=a\sqrt{1-e^2}\sin E,$$

其中 a,e 为圆锥曲线的半主径和偏心率.

开普勒方程(Kepler equation)　二体问题运动方程中的一个积分. 它给出天体在轨道上的位置与时间的函数关系. 对于椭圆轨道，开普勒方程为

$$E-e\sin E=n(t-t_0),$$

其中 e 为轨道偏心率，t 为时间，t_0 为天体过近日点时间，n 为平均角速度，$n(t-t_0)=M$ 称为平近点角，E 为辅助变量，称偏近点角，为天体于 t 时在轨道上的位置与近日点相对于椭圆轨道中心的角距，即从近日点起算的角度. 相对于日心的角距，则称为真近点角 f. 对于抛物线轨道，开普勒方程为

$$\tan\frac{f}{2}+\frac{1}{3}\tan^3\frac{f}{2}=\frac{\sqrt{2\mu}}{2}q^{-3/2}(t-t_0),$$

式中 f 为真近点角，$\mu=G(M+m)$，G 为万有引力常数，$M+m$ 为两天体质量之和，q 为轨道近日距. 对双曲线轨道，开普勒方程为

$$e\mathrm{sh}F-F=\nu(t-t_0),$$

式中 $\nu=\sqrt{\mu}a_1^{-3/2}$，a_1 为双曲线轨道半主径，F 为辅助变量，与向径 r 的关系为

$$r=a_1(e\mathrm{ch}F-1).$$

轨道计算(orbit determination)　亦称轨道确定. 确定天体轨道的方法之一. 它研究如何利用天体的观测资料和天体力学的有关定律确定天体初始轨道的方法. 多种轨道计算的方法可归结为数学问题，为两大类：一是根据观测资料计算天体在某一时刻的位置和速度，然后再由这时天体的位置和速度确定六个轨道根数. 实际上这是微分方程中的初值问题. 牛顿(Newton，I.)最早提出的图解法和 1780 年拉普拉斯(Laplace，P.-S.)提出的分析方法都属于这一类. 另一类方法是根据观测资料计算天体在两个时刻的位置，然后利用这两个位置确定六个轨道根数，这属于微分方程的边值问题，如计算小行星轨道的高斯方法和计算彗星轨道的奥伯斯方法. 由于采用了多普勒测速、雷达、激光等现代化的观测手段，对于某些天体，人们不仅能观测到它们的方向，也能测到它们的距离、速度. 在实践中，人们提出了

不少比经典方法更简便的轨道计算方法，它们多数用于确定人造天体的轨道.

轨道确定(orbit determination)　即"轨道计算".

初轨(preliminary orbit)　亦称无摄轨道.一种天体运行轨道.是求解二体问题得出的一种轨道.是用观测资料按二体问题理论算出的轨道.初轨的计算就称为轨道计算.在一些工程学科(如航天工程)中，用第一圈内的观测资料所计算的椭圆轨道(无论是否计算摄动)都称为初轨.

定轨(definitive orbit)　一种天体运行轨道.是经准确摄动计算和轨道改进后所得的轨道，是一种接近天体真实轨道的运行轨道.有时称为精确轨道或精轨，与初轨相对应.

拉普拉斯方法(Laplace method)　初轨计算的基本方法之一.是用三个时刻的行星或卫星在天球上位置的观测资料算出其椭圆轨道的方法.由拉普拉斯(Laplace，P.-S.)于1784年首先提出，后经很多人不断改进，一直沿用至今，但仍称为拉普拉斯方法.基本原理是：根据三个时刻的观测资料，用迭代法算出天体在第二时刻的日心(对行星而言)或地心(对人造卫星而言)的直角坐标(x,y,z)及其变化率

$$\frac{\mathrm{d}x}{\mathrm{d}t},\ \frac{\mathrm{d}y}{\mathrm{d}t},\ \frac{\mathrm{d}z}{\mathrm{d}t},$$

然后再用这六个量算出此时刻的六个轨道要素.

高斯方法(Gauss method)　初轨计算的基本方法之一.是用三个时刻的行星或卫星在天球上位置的观测资料算出其椭圆轨道的方法.是由高斯(Gauss，C.F.)计算新发现的小行星轨道时建立的，正式发表于1809年，后经很多人改进，至今还在应用，但仍称为高斯方法.基本原理是：根据三个时刻的观测资料，用迭代法算出天体在第一和第三时刻的日心(对行星而言)或地心(对人造卫星而言)的直角坐标($x(1),y(1),z(1)$)，($x(3),y(3),z(3)$)，然后再用这六个量算出六个轨道要素.

奥伯斯方法(Olbers method)　初轨计算的基本方法之一.是用三个时刻的彗星在天球位置的观测资料算出其抛物线轨道的方法.由奥伯斯(Olbers，H.W.M.)于1797年首先提出，其原理见"彗星轨道计算".

巴日诺夫方法(Bazhnov method)　初轨计算的基本方法之一.是用三个时刻的人造卫星在天球位置的观测资料算出其椭圆轨道的方法.由苏联的巴日诺夫(Бажнов，A.)于1958年提出.其原理是：根据卫星的三个位置在空间的几何关系，以及在轨道上应满足的朗贝特定理，用迭代法算出此三个位置到观测站距离和轨道面的方向，再由这些量算出椭圆轨道要素.若能从观测先定出较准的公转周期及相应的轨道半长径，用此法计算初轨就很有效.

朗贝特定理(Lambert theorem)　椭圆轨道上任意两个时刻的位置坐标同时间的关系定理.设天体沿椭圆轨道运动，在时刻$t(1)$，$t(2)$时的向径为$r(1)$，$r(2)$，此两位置之间的连线长度为s，则由二体问题公式可得下面关系：

$$n[t(2)-t(1)]=g-d-(\sin g-\sin d),\qquad(1)$$

其中n为椭圆运动的平均角速度；g，d为辅助量，与$r(1)$，$r(2)$，s有下面关系：

$$\sin\left(\frac{g}{2}\right)=\sqrt{\frac{r(1)+r(2)+s}{4a}},$$

$$\sin\left(\frac{d}{2}\right)=\sqrt{\frac{r(1)+r(2)+s}{4a}},$$

其中a为椭圆轨道半长径.(1)式就是朗贝特定理，有时又称朗贝特方程.

欧拉公式(Euler formula)　抛物线轨道上任意两个时刻的位置坐标与时间的关系的公式.设天体沿抛物线轨道运动，在时刻$t(1)$，$t(2)$时的向径为$r(1)$，$r(2)$，此两位置之间的连线长度为s，则由二体问题公式可得关系

$$6\sqrt{G(M+m)}[t(2)-t(1)]$$
$$=\sqrt{r(1)+r(2)+s}+\sqrt{r(1)+r(2)-s},$$

称此式为欧拉公式，有时又称为欧拉方程，其中G，M，m分别是万有引力常数和二体的质量.

彗星轨道计算(orbit determination of comet)　计算彗星轨道的方法.绝大多数彗星轨道接近抛物线，奥伯斯(Olbers，H.W.M.)于1797年提出一个完整的抛物线轨道计算方法.他利用力学定律导出抛物线轨道上的天体在两个时刻的向径和弦长应满足的欧拉方程，又利用太阳、地球和彗星三者应满足的几何约束条件，得到一个奥伯斯方程，用逐次逼近法求出彗星的空间位置，从而确定彗星的轨道要素.

轨道改进(orbit improvement)　对初轨的改进方法.初轨与天体的真实轨道相差甚大，需要用大量的观测资料进行修正，才能得到天体的精密轨道.目前，人们常用微分改进法对轨道进行修正.在某时刻，对天体进行观测，得到一个观测位置，用$\boldsymbol{r}_0$表示.在同一时刻，通过计算，得到一个理论位置，用$\boldsymbol{r}_c$表示，两者之差$\Delta\boldsymbol{r}=\boldsymbol{r}_0-\boldsymbol{r}_c$称为残差.如果认为消除了观测值的系统误差和偶然误差，那么残差就是理论值的误差，而理论值的误差主要是由于初始轨道的不精确产生的.因此，反过来通过多次观测，利用多次残差，用最小二乘法可算出轨道根数的改正值.用测速值$\dot{\rho}$和测距值ρ也可进行轨道改进；计算摄动后经轨道改进的初轨值又称为平均轨道根数.

平均轨道根数(mean orbit elements)　见"轨道改进".

小行星轨道计算(orbit determination of minor planet) 确定小行星轨道的方法. 1801 年，皮亚齐(Piazzi,G.)发现了第一号小行星(谷神星)不久，高斯(Gauss,C. F.)第一次算出了它的椭圆轨道，并于 1809 年正式发表. 高斯利用小行星轨道位于过太阳的一个平面内和运动过程中所遵循的力学定律，用逐次逼近法求出小行星向径在两个时刻内扫过的扇形面积与这两个时刻小行星位置与太阳构成的三角形面积之比值，然后求出小行星在直角坐标系的位置，最后由两个时刻的位置确定小行星的轨道. 高斯方法是用三次光学观测资料计算小行星轨道的有代表性的方法，后人对此方法有些改进，但基本原理没有什么变化.

引力作用范围(sphere of gravitational action) 天体力学术语. 指的是两个天体所形成的系统中质量较小的天体 m(其质量也用 m 表示)周围的一个区域，它的边界一般取以 m 为中心、以长度 r 为半径的球面，因此又称为引力作用球或影响球. 设质量较大的天体 M(其质量也用 M 表示)与 m 的距离为 d，一质点 P 在 m 的引力作用范围之内时，P 主要在 m 的引力作用下运动，而将 M 对 P 的引力作为摄动力；当 P 在 m 的引力作用范围之外时，P 主要在 M 的引力作用下运动，而以 m 对 P 的引力作为摄动力. 在近似讨论时，往往先忽略摄动力，将 P 在 m 和 M 的引力作用下的三体问题简化作两个二体问题. 引力作用范围 r 有三种不同的取法：

1. 以 m 和 M 对 P 点的直接引力为标准，m 的引力作用范围半径

$$r = r_1 = d\left(\frac{m}{M}\right)^{1/2},$$

r_1 又称引力范围.

2. 以对 P 点的引力和摄动力大小之比作标准，m 的引力作用范围半径

$$r = r_2 = d\left(\frac{m}{M}\right)^{2/5},$$

r_2 又称为用范围.

3. 以平面圆型限制性三体问题的拉格朗日特解 L_1 到小天体的距离作为引力作用范围半径，即

$$r = r_3 = d\left(\frac{m}{M}\right)^{1/3},$$

r_3 又称希尔范围.

三种数值相差很大，例如，在日地系统中，地球的三种引力作用范围分别为

$$r_1 = 2.7\times10^5\text{km},$$
$$r_2 = 9.3\times10^5\text{km},$$
$$r_3 = 2.27\times10^6\text{km};$$

在地月系统中，月球的引力作用范围分别为

$$r_1 = 4.3\times10^4\text{km},$$
$$r_2 = 6.6\times10^4\text{km},$$
$$r_3 = 8.9\times10^4\text{km}.$$

希尔范围(Hill gravitational action sphere) 见“引力作用范围”.

影响球(effective sphere) 即“引力作用范围”，常在航天工程中使用.

宇宙速度(cosmic velocity) 天体力学中常用的三种速度. 是从地面发射飞行器到宇宙空间所需的最小速度. 常用到的是下面三种：

1. 第一宇宙速度. 从地面发射的飞行器能成为人造卫星所需的最小速度，用 $V(1)$ 表示. 计算公式为

$$V(1)=\sqrt{\frac{GE}{a}}=7.905\text{km/s},$$

其中 GE 为万有引力常数与地球质量的乘积，又称地心引力常数(参见“引力常数”)，a 是地球赤道半径，$V(1)$ 也是地面环绕速度(参见“环绕速度”).

2. 第二宇宙速度. 从地面发射的飞行器能脱离地球引力场成为人造行星所需的最小速度，用 $V(2)$ 表示. 计算公式为

$$V(2)=\sqrt{\frac{2GE}{a}}=\sqrt{2}\,V(1)=11.18\text{km/s},$$

$V(2)$ 也是地面脱离速度(参见“逃逸速度”).

3. 第三宇宙速度. 从地面发射的飞行器能脱离太阳系引力场到恒星空间所需的最小速度，用 $V(3)$ 表示. 计算公式为

$$V(3)=\sqrt{V(2)^2+\frac{(\sqrt{2}-1)^2GS}{A}}$$
$$=16.65\text{km/s},$$

其中 A 为日地平均距离，即一个天文单位；GS 为万有引力常数与太阳质量的乘积，也就是日心引力常数(参见“引力常数”).

环绕速度(circular velocity) 天体力学中所研究的一种速度. 这里指圆形轨道速度. 与质量为 M 的某天体的距离为 R 的环绕速度 V 的计算公式为

$$V = \frac{GM}{R}.$$

逃逸速度(escape velocity) 天体力学中所研究的一种速度. 这里指一物体脱离某天体引力场束缚飞往星际空间所需要的速度，也称脱离速度. 因该物体的运行轨道为一抛物线，故又称抛物线速度. 逃逸速度

$$V_e = \sqrt{\frac{2GM}{r}},$$

M 为天体的质量，G 为引力常数，r 为物体距天体中心的距离. 当 M，r 分别为地球的质量和半径时，这样的逃逸速度称为第二宇宙速度，为 11.18km/s. 对

于太阳系其他天体，其表面逃逸速度是：水星 4.3km/s；金星 10.3km/s；火星 5.0km/s；木星 59.5km/s；土星 35.6km/s；天王星 21.4km/s；海王星 23.6km/s；冥王星 1.2km/s；月球 2.4km/s；太阳 617.7km/s.

双二体问题(double two-body problem) 一种二体问题. 指航天飞行器轨道设计中的二体问题. 例如，由地面向月球发射飞行器，飞行器在月球作用范围(相对地球)内运动时，只考虑月球的引力，是二体问题；飞行器在月球作用范围外运动时，只考虑地球的引力，也是一个二体问题，故飞行器从地面飞到月球过程中，由两个不同的二体问题组成，就是一种双二体问题.

最小能量轨道(orbit with minimum energy) 亦称为双切轨道. 行星际飞行器的一种轨道. 若飞行器在行星际空间的轨道是椭圆，既同地球轨道相切，也同目标行星的轨道相切，这种轨道所需能量最小，但飞行时间长，称这种轨道为最小能量轨道.

双切轨道(bitangential orbit) 即"最小能量轨道".

摄动理论(perturbation theory) 天体力学的主要内容和最早形成的分支. 是研究天体受摄动力作用后的运动状态的理论和方法. 在二体问题中，天体的运动轨道是一条不变的圆锥曲线(参见"二体问题")，一旦受到其他天体的引力或其他力学因素的干扰，天体的运动就会偏离原来的轨道，这种现象称为摄动. 受摄动天体的运动方程比二体问题的运动方程复杂得多，没有完全的分析解，可用分析方法和数值方法求其近似解. 分析方法又称普遍摄动，数值方法又称特殊摄动. 摄动理论不仅用来研究天体的运动，在理论物理和工程技术上也被广泛应用，通常称微扰理论.

普遍摄动(general perturbation) 见"摄动理论".

特殊摄动(special perturbation) 见"摄动理论".

微扰理论(perturbation theory) 即"摄动理论".

摄动(perturbation) 天体力学的基本概念之一. 指一个天体围绕另一天体按二体问题运动时，由于受到其他天体的吸引或受其他因素影响而对正常圆锥曲线轨道的偏离. 前种原因造成的偏离称为引力摄动，后者称为非引力摄动. 如太阳辐射压(和太阳风)的斥力和地球大气阻尼所引起的对彗星运动的影响就是非引力摄动. 按摄动项结构的形式，可分为摄动周期项(即摄动量要素中具有起伏特征的项)、摄动长期项(即摄动量总是朝同一方向变化)和摄动混合项(即以时间与周期函数的积表示的项). 研究精确地确定摄动的大小和变化规律的理论和方法称为摄动理论. 用另外可积系统的轨道(称为中间轨道)作基础，也可以讨论摄动.

摄动力(disturbing force) 天体力学的基本概念之一. 指产生摄动的作用力. 在天体力学中常见的摄动力有天体间相互引力、非球形引力、辐射压力、介质阻力等.

受摄运动(disturbed motion) 天体力学所研究的运动之一. 指有摄动力作用下的运动，与无摄运动对应. 在天体力学中，无摄运动一般指二体问题. 故由二体问题运动方程中加上摄动项后所描述的运动都是受摄运动.

强摄动(strong perturbation) 一种摄动. 指摄动项量级较大的情况. 受摄运动常用小参数方法求解，故摄动项量级与无摄项相比一般都小于 0.001；若大于 0.1，就称为强摄动，不能用小参数方法.

潮汐(tides) 一种天文现象. 因月球、太阳对地球各处引力不同所引起的水位、地壳、大气的周期性升降现象称为潮汐. 海洋水面的周期性升降称为海洋潮汐，地壳的周期性升降称为陆潮，也称固体潮，大气的则称为大气潮. 造成潮汐的力称引潮力或起潮力，它是月球或太阳对地球表面的引力与对地心的引力之差. 引潮力与距离的立方成反比，与月球或太阳的质量成正比. 由于日地距离为月地距离的 390 倍，所以月球引潮力为太阳引潮力的 2.25 倍. 在月球连续两次上中天的时间间隔(24h50min)内每天有两次涨潮和两次落潮(月潮)，日潮不能单独地观测到，它只能影响月潮的大小，在朔月和望月时，日潮和月潮同时发生，发生大潮. 在上弦和下弦时，月潮为涨潮，同时日潮却为落潮，这时发生小潮. 在引潮力作用下，海水内部、海水与海底、地球内部物质摩擦造成能量损耗称为潮汐摩擦. 潮汐摩擦可使地球自转速率缓慢降低，每世纪可使日长增长 1～2ms，还可引起地球自转角动量减少，同时使月球远离地球，使月球绕地球公转周期变长.

吻切轨道(osculating orbit) 天体摄动方程的一组解. 由常数变易法建立起来的摄动方程得到天体的六个轨道根数的一组解 $\sigma_i(t)$ (σ 代表轨道根数，$i=1,2,\cdots,6$). 在某一瞬时 t_0，将由 $\sigma_1(t_0)$，$\sigma_2(t_0)$，…，$\sigma_6(t_0)$六个轨道根数确定天体在该瞬时的轨道. 此时，天体既在实际轨道上，也在瞬时轨道上，瞬时轨道与实际轨道相切，故称这一瞬时轨道为吻切轨道. 如若从 t_0 时刻以后摄动力突然消失，天体将沿这一吻切轨道运动，此时又恢复为二体问题.

吻切椭圆(osculating ellipse) 一种吻切轨道. 吻切轨道为椭圆时，称为吻切椭圆(参见"吻切轨道").

受摄运动方程(disturbed equations of motion)

表示受摄运动的方程. 在天体力学中一般是常微分方程组. 主要形式有直角坐标的受摄运动方程、拉格朗日(行星运动)方程、牛顿方程、正则形式的受摄运动方程等. 这里只给出直角坐标的受摄运动方程. 设(x,y,z)为所讨论天体(质量为m)相对中心天体(质量为M)的直角坐标, 则受摄运动方程为

$$\frac{d^2x}{dt^2}+\frac{G(M+m)x}{r}=\frac{\partial R}{\partial x},$$
$$\frac{d^2y}{dt^2}+\frac{G(M+m)y}{r}=\frac{\partial R}{\partial y},$$
$$\frac{d^2z}{dt^2}+\frac{G(M+m)z}{r}=\frac{\partial R}{\partial z},$$

其中G为万有引力常数, R为摄动函数. 其余几种受摄运动方程参见本卷相应条目.

拉格朗日(行星运动)方程(Lagrange (planetary) equation)　一种行星受摄运动的方程. 拉格朗日(Lagrange, J.-L.)提出一种讨论行星运动的方法, 他以行星的椭圆轨道根数作为变量, 推导出的一组行星受摄运动方程的具体形式为

$$\frac{da}{dt}=\frac{2}{na}\frac{\partial R}{\partial M_0},$$
$$\frac{de}{dt}=\frac{1-e^2}{na^2e}\frac{\partial R}{\partial M_0}-\frac{\sqrt{1-e^2}}{na^2e}\frac{\partial R}{\partial \omega},$$
$$\frac{di}{dt}=\frac{1}{na^2\sqrt{1-e^2}}\cot i\,\frac{\partial R}{\partial \omega}-\frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial \Omega},$$
$$\frac{d\Omega}{dt}=\frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i},$$
$$\frac{d\omega}{dt}=\frac{\sqrt{1-e^2}}{na^2e}\frac{\partial R}{\partial e}-\frac{1}{na^2\sqrt{1-e^2}}\cot i\,\frac{\partial R}{\partial i},$$
$$\frac{dM_0}{dt}=-\frac{2}{na}\frac{\partial R}{\partial a}-\frac{1-e^2}{na^2e}\frac{\partial R}{\partial e},$$

式中R为摄动函数, a,e,i,Ω,ω,M_0分别是轨道半长径、偏心率、轨道倾角、升交点角距、近日点角距和历元时刻的平近点角, n为平均角速度.

拉格朗日括号(Lagrange bracket)　一种变量间的关系. 指两组变量之间的一种偏微商关系. 在一般力学和天体力学中经常用到. 设$2n$个变量$x(i)$, $y(i)$为两个量u,v的函数, 其中$i=1,2,\cdots,n$, 则定义u,v的拉格朗日括号为

$$[u;v]=\frac{\partial x(i)}{\partial u}\cdot\frac{\partial y(i)}{\partial v}-\frac{\partial y(i)}{\partial u}\cdot\frac{\partial x(i)}{\partial v}.$$

泊松括号(Poisson bracket)　变量间的一种关系. 指两组变量之间的一种偏微商关系. 在一般力学和天体力学中常用到. 设u,v为$2n$个变量$x(i)$, $y(i)$的函数, 其中$i=1,2,\cdots,n$, 则定义u,v的泊松括号为

$$[u;v]=\frac{\partial u}{\partial x(i)}\cdot\frac{\partial v}{\partial y(i)}-\frac{\partial v}{\partial x(i)}\cdot\frac{\partial u}{\partial y(i)}.$$

牛顿方程(Newton equation)　受摄运动方程的基本形式之一. 无论摄动力是否存在, 摄动函数都可使用. 设S,T,W为摄动加速度分别在向径方向、轨道面上与向径垂直的方向(运动方向为正)、轨道面法线方向(向北为正)上的投影, 以轨道要素a,e,i,Ω,ω,M为基本变量的受摄运动方程可表示为

$$\frac{da}{dt}=\frac{2}{n\sqrt{1-e^2}}\left[e\sin fS+\frac{p}{r}T\right],$$
$$\frac{de}{dt}=\frac{\sqrt{1-e^2}}{ne}[\sin fS+(\cos f+\cos E)T],$$
$$\frac{di}{dt}=\frac{r\cos(f+w)}{na^2\sqrt{1-e^2}}W,$$
$$\frac{d\Omega}{dt}=\frac{r\sin(f+w)}{na^2\sqrt{1-e^2}\sin i}W,$$
$$\frac{d\omega}{dt}=\frac{\sqrt{1-e^2}}{nae}\left[-\cos fS+\left(1+\frac{r}{p}\right)\sin f\,T-\frac{r\cot i\sin(f+w)}{na^2\sqrt{1-e^2}}W\right],$$
$$\frac{dM}{dt}=\frac{1-e^2}{nae}\cos f-\frac{2r}{na^2}S-\frac{1-e^2}{nae}\left(1+\frac{r}{p}\right)\sin fT,$$

其中n,f,E分别为平均角速度、真近点角、偏近点角.

正则共轭变量(canonical conjugate variable)　一种特殊变量. 满足正则方程组的两组变量, 称为相互正则共轭, 一起组成正则共轭变量.

正则方程组(canonical equations)　分析动力学和天体力学中常用的反对称形式的运动方程. 设p,q都是n维向量, 满足方程

$$\frac{dq}{dt}=\frac{H}{p}\ \text{和}\ \frac{dp}{dt}=-\frac{H}{q},$$

其中$H=H(p,q,t)$是p,q和时间t的一个数量函数. 上式就称为正则方程组或哈密顿正则方程组; p,q称为正则共轭变量. 这种方程组是由哈密顿(Hamilton, W. R.)于1835年首先提出, 式中H称为哈密顿函数. 哈密顿的贡献是经典力学发展的新里程碑, 后来称为哈密顿力学, 它对现代物理学和天体力学的发展都有重大作用.

20世纪80年代以后, 发现正则方程组在数值计算上有优越之处. 中国计算数学家冯康和国内外其他专家建立了相应的辛算法.

德洛内变量(Delaunay variable)　天体力学中常用的一种正则共轭变量. 是由法国德洛内(Delaunay, C. E.)于19世纪60年代研究月球运动时提出的. 记号为$L,G,H;l,g,h$. 它们同轨道要素有下面简单关系:

$$L=G(M+m)a,$$
$$G=L\sqrt{1-e^2},$$
$$H=G\cos i,$$
$$l=nt+M_0,$$
$$g=\omega,$$

$$h=\Omega,$$

其中 G 为万有引力常数，M,m 分别是中心天体和所讨论天体的质量，n 为平均角速度，l 是平近点角. 当 e,i 很小时，常用另一种形式的变量，称为推广的德洛内变量，记号为 $L',G',H';l',g',h'$. 同原德洛内变量的关系为

$$L'=L,$$
$$G'=G-L=-L(1-\sqrt{1-e^2}),$$
$$H'=H-G=-G(1-\cos i),$$
$$l'=l+g+h,$$
$$g'=g+h,$$
$$h'=h.$$

庞加莱变量(Poincaré variable) 天体力学中常用的正则共轭变量之一. 由法国庞加莱(Poincaré,(J.-)H.)于 19 世纪末讨论三体问题时提出，记号为 $L'',G'',H'';l'',g'',h''$，同推广的德洛内变量(参见“德洛内变量”)$L',G',H';l',g',h'$ 有下面关系：

$$L''=L',$$
$$G''=\sqrt{-2G'}\sin g',$$
$$H''=\sqrt{-2H'}\sin h',$$
$$l''=l',$$
$$g''=\sqrt{-2G'}\cos g',$$
$$h''=\sqrt{-2H'}\cos h'.$$

正则变换(canonical transformation) 一种特殊变换. 正则共轭变量之间的变换称为正则变换. 通过正则变换，有时可使正则方程组简化或降阶，故正则变换也成为正则方程组的原则解法之一(参见“变换理论”).

常数变易法(method of variation of constant) 研究天体受摄运动的一种方法. 二体问题的轨道在空间是一条不变的圆锥曲线(椭圆、抛物线、双曲线乃至直线的任何一种)，这一轨道可以用六个独立的常数来表述，通常用六个轨道根数. 一旦天体受到摄动，轨道根数将随时间变化，借助常微分方程中的常数变易思想，可以建立以轨道根数(或相应的其他六个独立常数)为变量的摄动方程，这种方法称为常数变易法. 在天体力学中常用到的以此方法得到的受摄运动方程有两种：一是用摄动函数表达的拉格朗日(行星运动)方程；另一种是摄动加速度分量表达的牛顿方程.

变换理论(theory of transformation) 天体力学中研究天体运动方程中变换变量的处理方法的理论. 该方法将运动方程中的变量进行变换以降低运动方程的阶次和消除碰撞奇点，使新的运动方程结构简单，易于求解. 它包括两方面的内容：

1. 正则变换. 哈密顿方程又称为正则方程，它具有对称性等一些优点. 若变换后的新方程仍保持正则形式，则称该变换为正则变换. 如 1916 年，柴培耳(Zeipel, H. von.)曾用正则变换寻找循环坐标的方法处理天体力学问题.

2. 正规化变换. 指消除质点组运动方程中碰撞奇点的变换. 它使新的坐标成为新的自变量的解析函数，便于理论分析，并往往给出运动方程解的具体表达方式. 如在平面圆型限制性三体问题中，蒂勒变换被用来求解双不动中心问题等.

摄动理论的小参数方法(method of small parameter in perturbation theory) 经典摄动理论的主要内容. 是求受摄运动方程分析解的基本方法. 由于无法求出受摄运动方程的严格通解，只能用级数展开求其近似解. 第一步是展开为一些小参数的幂级数. 以大行星运动为例，由于大行星质量与中心天体(太阳)质量之比小于千分之一，故大行星质量就是天然小参数. 若只考虑太阳和两个大行星(三体问题)，取太阳质量为单位，两行星质量分别为 m,m'，则行星的任一轨道要素 a，在两行星相互引力摄动下可展开为 m,m' 的幂级数：

$$a=a(0)+ma(1,0)+m'a(0,1)+ma(2,0)$$
$$+mm'a(1,1)+m'a(0,2)+\cdots,$$

其中一般项系数 $a(i,j)$ 内，i,j 分别表示此项 m,m' 的幂次，$i+j$ 就是此项的摄动阶数. 小参数方法是天体力学的奠基者们(参见“天体力学”)逐步建立的，但理论上系统化则由庞加莱(Poincaré,(J.-)H.)于 19 世纪末完成.

摄动函数(disturbing function) 表示摄动力大小的一个标量函数. 习惯上用字母 R 表示. 该函数的梯度 $\frac{\partial R}{\partial \boldsymbol{r}}$ 表示摄动加速度. 摄动力可分两类：一类是除主星体以外的其他天体的引力，这类摄动力为保守力(有势力)，因此可用摄动函数的梯度来表示其大小. 另一类为非保守力(耗散力)，如空间介质的阻力，它们不能用摄动函数来表示.

摄动函数的展开问题(problem of development of disturbing function) 天体力学研究的课题之一. 这里是指将摄动函数展开为时间或代表时间的其他变量的显函数的研究课题. 各种摄动理论最终都要对受摄天体的摄动方程积分，如果用分析方法积分这类方程，必须将摄动函数表示成时间或所选择的其他自变量的显函数. 摄动函数还是天体轨道根数的函数，对于太阳系的大行星，可利用小偏心率和小倾角进行展开，最后的形式是一个三角级数和幂级数的混合级数. 研究摄动函数展开问题，实质上就是如何改进展开式的收敛性问题. 遇到大偏心率、大倾角的情况(多数彗星和某些人造天体的轨道)，由于摄动函数的收敛性很差，甚至不收敛，因而不能

用分析方法进行研究，而采用数值方法.

一阶摄动(first order perturbation) 受摄运动方程的一种解. 是受摄运动方程精确到小参数一次幂的解(参见“摄动理论的小参数方法”). 即摄动函数展开式中的一次幂项. 求解时，要展开摄动函数为时间(或各种近点角)的显函数后，再积分得出.

长期摄动(secular perturbation) 天体力学术语. 指摄动量随时间成单调无限增加或减小的部分. 它反映了天体运动状态的长期变化趋势，尤其是轨道半长径和偏心率的长期变化直接影响天体运动的稳定性问题. 英国梅塞奇(Message,P. J.)于 1982 年证明在各阶摄动中太阳系大行星的半长径无长期变化.

周期摄动(periodic perturbation) 天体力学术语. 指摄动量中成周期变化的部分. 周期摄动又可分为短周期摄动和长周期摄动. 如果周期的长短接近椭圆轨道的周期，称这些项为短周期摄动，一般地，这类摄动项比较多，但振幅较小. 长周期项的振幅比短周期项的振幅往往要大几个量级，因而对天体运动的影响也更大. 在一些特殊的情况下，例如两个行星之间发生轨道共振，人造卫星的轨道倾角接近临界角等，在分析解中会出现周期非常长、振幅非常大的摄动(参见“共振问题”).

长周期摄动(long period perturbation) 天体力学术语. 指周期很长的周期摄动项. 在受摄运动方程的分析解中，周期项是三角函数，其角变量一般是时间 t 的线性函数. 有时 t 的系数很小，则相应的周期很长；又因通过积分，此系数会出现在分母上，成为振幅很大的项. 以行星运动的三体问题为例，一般周期项的形式为

$$A\ \sin\frac{[(pn+p'n')t+B]}{(pn+p'n')},$$

其中 n, n' 分别为两行星的平均公转角速度，p, p' 为任意整数. 当 n 与 n' 之比接近简单分数时，会有适当的整数 p, p'，使得 $pn+p'n'$ 为小量，此项就成为长周期项. 例如，木星和土星的相互摄动中，有 $-2n+5n'=3.99''/\mathrm{d}$，相应周期为 890a. 此项振幅约 $50'$，曾在 18 世纪木星的观测中反映出来，成为当时木星运动的难题.

高阶摄动(high order perturbation) 受摄运动方程的一种解. 是受摄运动方程精确到小参数二次或更高次幂的解(参见“摄动理论的小参数方法”). 即摄动函数展开式中的二次幂项或更高次幂项. 在行星运动中，一般只精确到二阶. 个别问题到三阶，但只取其中主要项(因三阶摄动项数量太大，很难全部求出). 其他问题也如此.

德洛内方法(Delaunay method) 摄动理论的基本方法之一. 该方法的特点是：用一系列的正则变换，使正则形式的受摄运动方程逐步简化，最后得到严格解或更精确的解. 此法是由法国德洛内(Delaunay,C. E.)于 19 世纪 60 年代研究月球运动时提出的；19 世纪末发展成为德洛内-柴培耳方法. 人造卫星上天后，在解决卫星运动中起了重要作用，并有新的发展. 这样的正则变换已成为解正则方程组的一种原则方法.

汉森方法(Hansen method) 摄动理论的基本方法之一. 其特点有两个：一是提出理想坐标，把问题集中到只要计算三个摄动量；二是把摄动函数展开为偏近点角的三角级数，而且其中关键展开式采用了数值方法(后来就称为半分析方法). 这是汉森(Hansen,P. A.)于 19 世纪中研究小行星时提出的，并得到广泛应用.

希尔方法(Hill method) 摄动理论的基本方法之一. 其特点是：用近似运动方程的一条周期轨道(比椭圆更接近实际运动轨道)作为中间轨道，再进一步讨论摄动. 这样的方法后来就称为中间轨道法. 是由美国的希尔(Hill,G. W.)于 19 世纪末研究月球运动时提出的，后经美国布朗(Brown,E.)改进成为一种实用的月球运动理论，被采用作为天文年历计算月球历表的根据，并一直用到 1984 年.

中间轨道(intermediate orbit) 天体力学中研究的一种天体运动轨道. 是在天体力学摄动理论中一种假设的、接近于天体真实运动的、能列出解析表达式的轨道. 在通常的摄动理论中，以二体问题(作用力仅包含中心天体的引力)中的开普勒椭圆作为第一次近似，然后用迭代过程求各阶摄动，而求出一定精度的解. 但二体问题往往与真实情况相差很远，需要寻求一种比椭圆轨道更接近真实运动的解——中间轨道，即寻求一种作用力，不仅包含中心天体的引力，而且还尽可能多地包含一部分摄动力，对应的运动方程一般是可积的，进一步求摄动时要简单的可积系统. 例如，在限制性三体问题中，如果一个主天体绕另一个主天体的运动速度比小天体的运动速度小得多时，可以近似地看做是不动的，即小天体在两个固定的主天体吸引下运动，称为双不动中心问题，其对应解可以看做小天体运动的中间轨道. 希尔(Hill,G. W.)在月球运动理论中，用近似模型得到的解，称为二均轨道，将其作为中间轨道，再求摄动，是天体力学中用中间轨道的成功例子.

坐标摄动法(method of coordinate perturbation) 研究摄动运动的一种方法. 指以天体在实际轨道上的位置与在二体问题轨道(或其他中间轨道)上的位置差为基本变量来研究摄动运动的一种方法. 在历史上，直角坐标的摄动法、球坐标和极坐标的摄动法都曾被一些天文学家用来研究彗星、行星、月球以及各种人造天体的运动.

小行星的运动(minor planet motion) 天体力学中的一个研究课题. 小行星是太阳系的一类小天体,绕太阳沿椭圆轨道运动,多数集中在火星和木星轨道之间. 由于小行星的轨道偏心率和倾角普遍比大行星的大,所以一般地,用来研究大行星的分析方法不适于小行星. 历史上曾有汉森(Hansen, P. A.)、希尔(Hill, G. W.)等提出一些针对小行星轨道特点的分析方法,但目前最有效的、也是被广泛采用的是用数值方法解小行星的运动方程,直接得到在几十年甚至几百年内小行星的位置和速度.

汉森-波林方法(Hansen-Bohlin method) 小行星摄动计算的重要方法之一. 19 世纪后期,由于长时间露光照相定位方法广泛用于天文观测,小行星被大量发现,最多的一年就发现一百多颗小行星. 但因小行星很暗,如不能尽快计算摄动而预报较准位置,一两年后就会丢失. 丹麦波林(Bohlin, K.)以汉森方法为基础,提出一种所谓"群法",即以公转周期相近的小行星组成小行星群(参见"小行星群"),对同一群的小行星,只需用同一个摄动计算公式;而对不同的小行星只需作简单改正. 其精度能满足编制星历表的要求,被命名为汉森-波林方法.

共振问题(resonance problem) 天体力学中的一类研究课题. 即研究摄动量级数解中共振奇点的问题. 对于自然天体,在级数解的周期项振幅中会出现 $1/(pn-qn')$ 因子,n,n' 分别为被摄动天体和摄动天体的平均角速度,p 和 q 为正整数,当

$$\frac{n}{n'}=\frac{q}{p}$$

时,$pn-qn'=0$,出现共振奇点,就是所谓通约问题,而太阳系天体几乎都满足通约条件. 人造卫星带谐项摄动长期项振幅会出现 $1/(4-5\sin^2 i)$ 因子,当倾角 $i=i_c=63°26'$ 或 $116°34'$ 时,$4-5\sin^2 i=0$. i_c 称为临界角. 这二者只意味着级数解失效,不能说明轨道要素真的会变为无穷大(因为运动方程本身并无这样的奇点,上述奇点是所用方法带来的),只要方法作些改变就不会出现这个问题.

小行星环的空隙(gap of asteroid ring) 一种天文现象. 指小行星数目分布的一种不均匀现象. 若用 n,n_1 分别表示小行星和木星的平均角速度,根据开普勒第三定律,平均角速度与轨道半长径有关(参见"开普勒第三定律"),因此不同的平均角速度代表大小不同的轨道. 目前已发现当 n/n_1 满足 1/2, 2/5, 1/3时,小行星分布出现空隙;而当 n/n_1 满足 1, 3/4, 2/3 时,小行星分布又出现密集.

小行星群(group of asteroids) 小行星的组合. 这里是指公转周期相近的小行星组合. 为了与共振问题联系,以小行星周期与木星公转周期之比值 P/P' 为准,此比值接近某简单分数时分为一群,并命名如下表. 各群中心区有的是空隙,有的是小行星聚积(参见"小行星环的空隙"和"共振问题").

小行星群名称和相应的 P/P' 所接近的简单分数

P/P'	1	3/4	1/2	2/5
名称	Trojan	Hestia	Hecuba	Minerva
P/P'	1/3	2/7	1/4	
名称	Hilda	Flora	Hungary	

小行星族(family of asteroids) 小行星的组合. 这里是指轨道很接近的小行星组合. 20 世纪 20 年代,由日本的平山清次首先提出. 同一族的小行星几乎就在同一条轨道上,它们的轨道要素 a, e, i 都很接近. 至今已找出一百多个小行星族,并且正在研究各小行星族内每个星的物理性质(如反照率、化学成分等),试图表明同一族的小行星有同样的起源.

彗星的运动(motion of comet) 天体力学中的一个研究课题. 绝大多数彗星的运动轨道是接近抛物线的双曲线. 少数呈很扁的椭圆,这类彗星称为周期彗星. 周期彗星的轨道周期短的为几年、几十年,长的可达几千年甚至上万年,这种轨道很难与双曲线轨道相区别. 彗星运动的另一特点是它绕太阳运行的公转方向不一定都是由西向东的顺行(所有行星都顺行),可以是南北方向的运动,甚至是由东向西的逆行. 彗星椭圆轨道的近日距与远日距可相差很大,所以它与太阳的距离可由千分之几个天文单位变化到近百个天文单位. 多数彗星受木星的摄动很大,有些彗星受木星摄动后可完全改变原来的轨道,以致造成受摄前后似乎不是同一颗彗星的错觉. 从彗星的以上运动特点可以看出,用分析方法研究彗星的运动困难太大,因此,无论在历史上还是在目前,人们都采用数值方法研究彗星的运动.

天体力学定性理论(qualitative theory in celestial mechanics) 天体力学的较新的一个分支. 研究内容是:天体的运动和形状在长时期(包括时间趋于无穷)内的特性,以及轨道在运动方程奇点附近的性质. 天体力学诞生后的 200 年内,主要用分析方法求出天体运动方程的近似解. 由于所得近似解为级数形式,一般不收敛或收敛很慢,无法用于讨论长时间内的运动性态. 1892 年,庞加莱(Poincaré, (J.-)H.)在他的名著《天体力学新方法》中,正式提出了定性方法,书中具体讨论了运动方程的可积性、周期轨道及相应的特征指数理论等,由此诞生了天体力学定性理论. 同时,俄国人李亚普诺夫(Ляпунов, A. M.)的著作《关于运动稳定性的一般问题》,使运动稳定性理论成为天体力学定性理论的主要内容之一. 他们又进一步讨论了天体(看成流体)的平衡形状和稳

定性问题，成为天体力学定性理论的新领域. 1927年，数学家伯克霍夫(Birkhoff, G. D.)的名著《动力系统》出版后，天体力学定性理论同它相结合，有更快发展，成为一种特殊的动力系统. 以后随着拓扑学和微分方程定性理论的发展，不断用于天体力学定性理论，深化了所研究的内容，故有些文献中称天体力学定性理论为“天体力学的拓扑方法”. 20 世纪 60 年代以后，卡姆理论的建立和用数值方法讨论定性课题的出现，促使天体力学定性理论的发展形成一个新高潮. 不仅丰富和发展了运动稳定性理论的研究领域，还建立了新领域“定性理论的数值探索”. 由于电子计算机水平的迅速提高，天体运动的短时间内运动性态和位置预报等大量的定量课题都可用数值方法直接算出，而且随着各种计算软件的定型和规格化，有关研究工作量会不断减小，因此定性理论在天体力学整个学科中的份量会不断增加. 现代天体力学定性理论主要有下列内容：

1. 研究天体在奇点(运动方程奇点或级数解奇点)附近的运动性态，如碰撞奇点附近产生的俘获和交换、轨道共振、流形拓扑结构等.

2. 研究时间趋于无穷时的天体运动性态，如运动稳定性理论、卡姆理论、周期轨道理论等.

3. 定性研究的数值探索，即用数值计算来探索预期的结果.

此外，一些特殊定性课题也形成专门的研究领域，如三体问题的定性研究、太阳系稳定性问题等.

天体运动稳定性理论(theory on stability of the motion of celestial body)　天体力学定性理论的重要研究领域. 它研究各种类型天体运动在时间趋于无穷时的特性. 随着微分方程定性理论和动力系统的发展，研究各种天体运动的稳定性理论也不断发展. 在很多有关运动稳定性的理论中，天体力学主要用到下列几种：

1. 李亚普诺夫稳定性. 又分为两种：

1) 初值稳定性，即初值相差足够小时，相应天体位置相差永远很小. 设 $r(t)$为天体轨道，对任一正数 ε，存在正数 $\delta=\delta(\varepsilon,t_0)$，使得邻近轨道 $r'(t)$，在 $|r'(t_0)-r(t_0)|<\delta$ 时，有

$$|r'(t)-r(t)|<\varepsilon,$$

则称 $r(t)$为稳定的. 这种稳定性在自然天体运动中用得不多，而发射人造天体时常用，可推广到相空间.

2) 摄动稳定性. 设无摄运动时的解为 $r(t)$，运动方程增加摄动项 $R(r,t)$后，相应的解为 $r'(t)$，对任意给定的正数 ε 和初始时刻 t_0，存在正数 $\delta=\delta(\varepsilon,t_0)$，使得 $|R(r,t)|<\delta$ 时，有

$$|r'(t)-r(t)|<\varepsilon \quad (\text{对所有 } t>t_0),$$

则称 $r(t)$为稳定的. 这种稳定性多用于特殊天体轨道(如平动解).

2. 轨道稳定性. 设天体轨道 $r(t)$在 $t_0<t<\infty$ 段的轨线为 L^+，并记某点 P 与 L^+ 的最短距离为 $\rho(P,L^+)$，若任意给定正数 ε，存在 $\delta=\delta(\varepsilon,t_0)>0$，使得 $|r'(t_0)-r(t_0)|<\delta$ 时，有

$$\rho(r'(t),L^+)<\varepsilon,$$

则称 $r(t)$为轨道稳定. 这种稳定性在天体力学中有广泛应用，表明天体的轨道变化很小.

3. 拉格朗日稳定性，又称为区域稳定性. 设 R^n 为天体的位形空间，$G^n\subset R^n$ 为一有限子集，即 G^n 中任意两点的距离小于某给定正数. 若天体运动方程的解为 $x(t)$，而 $t_0<t<\infty$，在 $x(t_0)\in G^n$ 时，有

$$x(t)\subset G^n \quad (t>t_0),$$

则称此解在拉格朗日意义下稳定. 这种稳定性多用于天体系统(如星团、星系等)的动力稳定性.

4. 希尔稳定性. 来源于圆型限制性三体问题的希尔曲面，或零速度面. 设在两个有限体 P_1，P_2 中，P_2 的质量较小，当初始条件使得相应的雅可比积分常数 C 的值满足 $C_1<C$ 时，零速度面有一分支是围绕 P_2 的闭曲面. 只要质量为无穷小的天体初始条件使得相应的 $C>C_1$，而且位置在此闭曲面内，则可证明它永远不会穿出此闭曲面. 这也可以看成一种区域稳定性，称为希尔意义下的稳定性. 在天文学中有广泛应用，特别是用来讨论双星运动. 现在已证明一般三体问题也有相应的希尔曲面，故可推广到一般三体问题情况.

一些专门的天体系统，如太阳系等，它们的稳定性已成为专门课题. 近年来也用数值方法探索天体运动的稳定性，其中用得较多的是李亚普诺夫特征数(LCN).

李亚普诺夫稳定性(stability of Liapunov)　运动稳定性的标准之一. 若天体初始条件(初始时刻的坐标和速度分量或轨道要素)的误差在某一小范围内，当时间充分增大时，天体坐标和速度分量的偏差永远局限在一定的范围内，这就是天体运动在李亚普诺夫意义下稳定的概念，是俄国的李亚普诺夫(Ляпунов, A. M.)在 19 世纪首先提出的. 此稳定性在天体运动中很少用(因一般的运动轨道都不满足，例如二体问题的椭圆轨道就不满足).

轨道稳定性(orbital stability)　运动稳定性的标准之一. 若天体在初始时刻的轨道要素有小误差，当时间不断增大时，轨道要素的偏差永远局限在一定的范围内，这就是轨道稳定性概念. 在天体运动理论中常用到此种稳定性.

希尔稳定性(stability of Hill)　运动稳定性的标准之一. 若天体于初始时刻在希尔范围内，当时间不断增大时，天体永远保持在希尔范围内(参见“引

力作用范围”)运动,则称此天体运动满足希尔稳定性条件.

稳定区(stable region)　天体力学术语.用数值探索算出不同初值得到的映射图象有不同结果:有孤立点、闭曲线和无秩序的弥散状态三种情况.前两种情况对应的李亚普诺夫特征数 LCN≤0;后一种情况的 LCN>0.使 LCN≤0 的初值所在区域称为稳定区或有序区;使 LCN>0 的区域为混沌区.有些动力系统还会出现中间状态,称为半有序区.

太阳系的稳定性问题(problem on the stability of solar system)　天体力学中的一个基本问题.研究太阳系在自身引力作用下是否能保持现在的状态,即各行星顺序不变,轨道形状和大小大体接近现在情况;有无大行星逃逸或碰撞甚至堕落到太阳上.这是天体力学定性理论和天体演化学的一个基本问题.虽然已研究了三百年,但至今仍未解决.牛顿时代虽有研究,但未给出具体结果.第一个用理论给出太阳系稳定性结论的是 18 世纪的拉普拉斯(Laplace,P.-S.)和拉格朗日(Lagrange,J.-L.),这就是后来以他们二人名字命名的定理,用公式表示为

$$\sum_{i=1}^{n} m_i \sqrt{a_i} e_i^2 = \text{常数} + \text{周期项},$$

$$\sum_{i=1}^{n} m_i \sqrt{a_i} \left(\tan \frac{I_i}{2}\right)^2 = \text{常数} + \text{周期项},$$

式中 m_i, a_i, e_i, I_i 分别为第 i 个大行星的质量、轨道半长径、偏心率、轨道倾角,n 为大行星个数.此二式表明太阳系是稳定的.但他们的这个结果,只考虑了行星间引力的一阶摄动,只能是粗略结论.在二阶摄动基础上,泊松(Poisson,S.-D.)得到一个定理:大行星轨道半长径没有长期变化,即不会随时间增加而无限增大或缩小.这只是从一个侧面说明太阳系可能是稳定的.在高阶摄动基础上,不少人曾得出行星轨道半长径有长期变化的结果.但到 1982 年,梅塞奇(Message,P.J.)终于证明了在任意阶摄动下,大行星轨道半长径仍没有长期变化.尽管这个定理很重要,但也只能说明太阳系可能是稳定的,因为若某行星轨道偏心率如增加到大于 1,仍会出现逃逸而不稳定.卡姆理论出现后,用于研究太阳系这个动力系统,从概率意义上得到有关太阳系稳定性的一个定理,称为阿诺尔德定理:若太阳系各行星质量(以太阳质量为 1)、轨道偏心率、轨道倾角都是小量,则在行星相互引力作用下,各行星的运动几乎永远是拟周期轨道,即保持在很细的环面上.定理中的所谓“几乎”,表示不是拟周期轨道的概率为 0,也就是说可认为太阳系几乎是稳定的.当然,行星质量等于小量,是否符合太阳系实际,还很难判定,因此,还不能最后肯定太阳系是稳定的.近年来由于计算机水平不断提高,用数值方法研究太阳系的稳定性成为可能,形成了太阳系稳定性的数值探索这个新领域.现在已有五个研究组在进行探索,其中美国两个,英、法、日各一个.一般都是采用太阳系质心中大行星运动的平均化方程,用数值方法计算各大行星轨道根数 a_i, e_i, I_i 的变化,若没有长期变化或振幅很大的长周期变化,则判定太阳系为稳定的.现在都已算到几千万甚至 1 亿年的时间范围,看来结果都表明太阳系为稳定的.但法国经度局的一研究组计算各大行星运动的李亚普诺夫特征数(LCN)接近于 0.2,表明太阳系在时间趋于无穷时为混沌状态,即为不稳定情况.当然这都还不是最后结论,有待进一步研究.

阿诺尔德定理(Arnold theorem)　见“太阳系的稳定性问题”.

卡姆理论(KAM theory)　天体力学中的一种稳定性理论.指一种关于哈密顿正则方程组的解的稳定性理论.由柯尔莫哥洛夫(Колмогоров,А.Н.)于 1954 年提出原理,1962 年,再由阿诺尔德(Арнольд,В.И.)和莫泽(Moser,J.K.)解决共振奇点和解析条件等困难而完成,后以他们三人姓氏第一字母命名为 KAM 理论.他们证明了哈密顿正则方程组在小摄动下的解为拟周期轨道(保持在某个很细的环面内),不稳定的概率为 0.并用它解决了圆型限制性三体问题的三角平动解的稳定性问题,还得到了有关太阳系稳定性的一个结果(参见“太阳系的稳定性问题”).

周期轨道理论(theory of periodic orbit)　天体力学中有关轨道的理论.指研究天体运动中周期轨道的存在性和稳定性的理论.由于天体运动方程一般不能严格解出,有必要在不用直接解而寻找出一些有用的特殊轨道,周期轨道就是最重要的一种.周期轨道理论是庞加莱(Poincaré,(J.-)H.)于 1892 年首先创立的较系统的理论.庞加莱在《天体力学新方法》第一卷中,用小参数方法证明了三体问题有三种类型的周期轨道存在,后来命名为庞加莱第一、二、三类周期轨道.后经莫尔顿(Moulton,F.R.)推广到普通直角坐标系统,又由达尔文(Darwin,G.H.)和施瓦茨席尔德(Schwarzschild,K.)结合具体天体情况求出几种周期轨道,使周期轨道理论得到肯定和发展.由于周期轨道的特点,经一定时间后,天体位置和速度回到原来数值,故在相空间中,周期轨道在某种拓扑映射下是不动点.随着拓扑学的不断发展,用拓扑方法研究周期轨道的存在性成为天体力学的重要内容,后来伯克霍夫(Birkhoff,G.D.)及阿诺尔德(Арнольд,В.И.)等人将动力系统方面的成果用于周期轨道研究,使周期轨道理论逐渐系统化,成为天体力学定性理论中的活跃领域.在天体

的实际运动中，严格周期轨道是不存在的，因此，研究与周期轨道有小偏差(包括初值小偏差和运动方程增加小摄动)的轨道特性更有实用意义. 由此形成两个方向：一是研究周期轨道的稳定性，即讨论偏差在时间无限增大时是否保持在小范围内. 庞加莱提出的特征指数理论就是讨论周期轨道稳定性的方法. 因只讨论偏差微分方程的线性项，所得的稳定性又称为线性稳定. 近年来对高阶稳定性的探讨中形成了不少方法，但尚未系统化. 另一个是中间轨道理论，即以所得的周期轨道作为无摄运动的基本轨道，在天体力学中又称为中间轨道，然后再讨论中间轨道的摄动，得到实用的轨道. 希尔(Hill, G. W.)的月球运动理论就是用它求出的一种周期轨道，命名为二均轨道，将它作为中间轨道，再以它为基础讨论摄动，是天体力学中的一个成功例子.

周期轨道理论除用定性方法研究外，也用分析方法和数值方法. 庞加莱在 1892 年提出周期轨道理论时，就用分析方法把周期轨道表示为小参数的幂级数，并用逐次积分求出此级数的系数. 但在实际问题中，积分过程很繁，很难求到幂级数的高阶项系数，而且收敛性证明也很困难. 20 世纪 60 年代以后，由于电子计算机广泛应用，产生了用数值方法求周期轨道的途径，而且在一段时期形成了高潮. 但数值方法也与分析方法一样，对周期长的轨道进行讨论比较困难，只能作为定性方法的补充.

定性理论的数值探索(numerical exploration of qualitative theory)　天体力学定性理论的一个研究领域. 它用数值方法检验和发现天体运动的一些定性结果，也是动力系统数值探索的一个领域. 20 世纪 60 年代以后正式形成. 1964 年，厄农(Henon, M.)等第一次用庞加莱截面法对具有轴对称引力场的星系进行研究，用映射的数值计算探索第三积分的存在性. 这项工作的结果标志着定性理论的数值探索正式诞生. 他们的工作受到天体力学、理论物理学、应用数学专家们的高度重视，发展很快，现已应用到化学、生物学、气象学以及一些社会学课题. 在天体力学中，由天体运动方程所决定的动力系统是否可积，是一个重要的定性课题. 用截面法便于讨论两个自由度的保守系统，因为可把此四维动力系统化为二维截面上与轨线交点的映射，容易绘图. 周期轨道在截面上为一个或有限个孤立点，拟周期轨道是一条或有限条闭曲线，故若截面上交点映射图象只有孤立点或闭曲线，则原动力系统为可积，相应的孤立点为不动点，闭曲线为不变曲线. 若映射图象为弥漫状无秩序的密集点，即为混沌状态，对应不可积系统. 厄农的例子表明，一些初值为不变曲线，另一些初值对应混沌，这表明原动力系统不是全局可积. 1966 年，希腊一个研究组用于讨论平面圆型限制性三体问题，有类似结论. 近年来有很多动力系统用这种方法进行了研究. 对于更高维的保守系统，也可用这种方法化成二维的映射，只是绘图困难.

在一般情况下，截面上映射图象中的一些区域呈现孤立点和不变曲线，称为有序区，另一些区域呈现混沌状态，称为混沌区. 1979 年，意大利一个研究组发现，有序区内邻近两点的距离随迭代次数(时间)线性增长，而混沌区则按指数函数形式增长. 他们提出用初始点偏差距离的指数增长率，又名李亚普诺夫特征数(简称 LCN)来判断有序区或混沌区. 混沌区的 LCN 值大于 0，有序区的 LCN 值等于 0. 1980 年以后，逐步建立了有关 LCN 的理论和计算方法，用于研究大量动力系统. 从 LCN 的定义可看出，还能用它判断出相应初始点出发的轨道是否稳定：LCN 值大于 0，为不稳定；LCN 小于 0，为稳定，而且是渐近稳定；LCN 值等于 0 时，为中间状态，但一般认为属于稳定范畴，称中性稳定. 这样，就讨论了天体力学定性理论的另一重要领域——运动稳定性理论. 现在已有广泛应用. 大量的数值探索表明：天体运动在很长时间后会出现混沌状态. 这样，天体力学被认为是在对确定论代表的传统观念进行挑战；同时，也把长期对立的天体力学和统计力学结合起来，将会影响天体力学的未来发展方向.

李亚普诺夫特征数(Liapunov characteristic number)　简称 LCN. 动力系统随机性及轨道稳定性的一种指示量. 最初由李亚普诺夫(Ляпунов, A. M.)在讨论微分方程组的解的性质时提出，现代用于一般动力系统. 这里只给初值偏差的指数增长率一个通俗定义：设时间 $t=0$ 时初值 x_0 与 $\overline{x}_0$ 的距离 $|x_0-\overline{x}_0|=d_0$，它们各自沿自己轨线变化，到任意时刻 t 时的值为 $x(t)$ 和 $\overline{x}(t)$，记

$$d=|x(t)-\overline{x}(t)|=d_0e^{\lambda t},$$

定义

$$\lambda=\lim_{t\to\infty}\frac{1}{t}\ln\frac{d}{d_0},$$

称 λ 为动力系统在 $\overline{x}_0$ 处的 LCN 值，即李亚普诺夫特征数. $\lambda>0$ 时，d 为指数增长；$\lambda<0$ 时，d 为指数减小；$\lambda=0$ 时，d 对应非指数函数增长(如 t 的多项式).

严格定义要用到流形上切空间的微分与胚切映射概念.

碰撞奇点(singular point of collision)　运动方程的主要奇点之一. 由于万有引力与距离的平方成反比，故天体运动方程中有距离在分母上，距离等于零(即碰撞)时为运动方程奇点，称为碰撞奇点. 二体碰撞奇点为可去奇点，用正规化变换能够消掉.

俘获问题(capture problem)　三体问题动力演化现象之一. 若三体 A,B,C 中，A,B 两体相对运动

为椭圆，C 体相对 A,B 的质心运动轨道为双曲线. 当 C 体从无穷远处到达 A,B 两体附近时，由于相互引力作用，可能使 C 体相对于 A,B 两体质心的轨道变为椭圆；而 A,B 两体相对运动仍保持为椭圆轨道，则称 C 体被 A,B 两体俘获. A,B 两体可推广到多体系统，有可能使 C 体被俘获. 太阳系中彗星从无穷远处来，受木星摄动后被俘获为椭圆轨道的情况不少.

交换问题(exchange problem) 三体问题动力演化现象之一. 若 A,B 两体相对运动为椭圆轨道，C 体相对 A,B 两体质心运动为双曲线轨道. 当 C 体从无穷远处到达 A,B 两体附近时，由于相互引力作用使 C 同 A 相对为椭圆运动，而 B 相对于 A,C 两体质心的运动变成双曲线轨道. 这相当于 B,C 体的位置交换，这种现象在天体力学定性理论中称为交换问题.

动力演化(dynamical cosmogony) 天体演化学和天体力学的共同研究领域. 即用动力学理论研究天体系统的演化问题. 虽然天体和天体系统的演化依赖于很多物理因素，但在漫长的演化过程中，总有相当长的历史阶段是以力学作用为主，故动力演化在天体演化学中有重要地位. 长期以来，动力演化也形成了一些重大课题，例如，太阳系的稳定性，引力 N 体问题的数值模拟、俘获和交换，轨道稳定性等.

天体力学数值方法(numerical method of celestial mechanics) 天体力学的三个基本方法之一. 是应用微分方程数值解法求解天体运动方程的方法，常称为特殊摄动方法，与分析方法、定性方法并列为天体力学三个基本方法. 数值方法早期是在传统的分析方法很难用于彗星和小行星运动后发展起来的. 数值方法比分析方法应用范围广，公式简单，精度高. 由于高速电子计算机的使用，对花费大量机时的矛盾有所缓和. 误差、稳定性和计算工作量是衡量数值方法的重要标准. 误差愈小、稳定性愈强、计算量愈小，则方法愈好，三者兼顾是很难的. 最早用于研究哈雷彗星运动的数值方法是科威耳方法，以天体直角坐标为变量，考虑天体在其他天体作用下的运动，运动方程

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=F(t,x)$$

是二阶的，右端函数不含$\frac{\mathrm{d}x}{\mathrm{d}t}$，这是多体问题数值解的传统的主要方法. 另一个以摄动量为变量的恩克方法用于计算短周期彗星和月球火箭的运动，也是个传统的常用的方法.

在星数 N 很大的天体系统中，N 体数值模拟方法有直接积分方法、差分网格法和蛙跳方法等，是广为使用的数值方法. 这些方法尤其适用于星团、星系、星系团这样庞大质点数的天体系统结构和演化的研究.

科威耳方法(Cowell method) 求解多体问题的一种数值方法. 20 世纪初，科威耳(Cowell，P. H.)研究哈雷彗星运动时，提出一种外推法，后来和克洛梅林(Crommelin，A. C.)进一步完善形成的数值方法，即为科威耳方法. 该方法采用直角坐标为变量，研究天体在其他天体的引力作用下的运动. 方程形式为

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=F(t,x),$$

它的右端函数 F 不含$\frac{\mathrm{d}x}{\mathrm{d}t}$，自变量为时间 t，通过对变量 $x=f(x)$ 作泰勒展开及求各次差分，求得变量不同时刻 t 的数值. 它是求解行星和彗星问题中的重要数值方法.

蛙跳方法(leap frog scheme) 解牛顿运动方程的一种数值方法. 此方法多用在无碰撞的大质点数的天体系统中，在差分网格法求得引力势后使用. 方法简单，程序简捷，占用机时少，精度是时步的平方$(\Delta t)^2$. 当它的误差传播方程的特征解位于单位圆内，即$|\lambda|\leqslant 1$，方法是稳定的. 在星系、星团的模拟研究中广泛使用，是个时间可逆的数值方法.

直接积分方法(direct integrational scheme) 一种天体力学数值解法. 是在引力 N 体问题模拟中，直接求解牛顿运动方程的一种数值方法. 它是将质点坐标和速度展开到四阶的泰勒级数多项式，与四阶均差多项式联合使用，由初始位置和初始速度，计算下一时刻质点的位置和速度. 此方法直接从运动方程出发迭代求解，计算量较大，一般只用来作星团的模拟研究.

差分网格法(particle-mesh method) 求解偏微分方程的一种数值方法. 在 N 体问题模拟中，网格(分二维或三维，直角坐标或极坐标)用来表示质点系统的质量分布和引力势分布. 由质点系统的质量分布，用差分网格法解泊松方程，得到质点系统在网格上的引力势分布的一种数值模拟方法，即为差分网格法. 求解引力势是 N 体问题模拟计算中一个重要环节，得到引力势，通过微分内插，就知道牛顿运动方程的右端项，接着解运动方程. 这是 N 体问题模拟计算中常用的方法.

引力N体问题计算机模拟(computer simulation of gravitational N-body problem) 天体力学中的一个研究领域. 是用数值模拟方法研究引力 N 体问题. 引力 N 体问题主要是用分析方法研究天体系统 N 个质点在万有引力作用下的运动规律，质点数并不多. 计算机模拟则是用数值方法研究像星团、

星系这样含有庞大质点数的恒星系统的结构和演化.质点一般采用等质量,初始位置和初始速度结合实际研究对象用伪随机数方法产生.当质点数 N 多达 10^5 量级,对于三维情况,联立微分方程多达 6×10^5量级.对于如此大的质点系统,一般采用差分网格法求解引力势,用时间中心的蛙跳方法求解微分方程,获得质点新的坐标和速度,再作为质点新的初值,如此连续迭代求解.此外还有直线积分法、球谐函数展开法、树形法、限制性三体问题等方法用于 N 体问题数值模拟中的引力势的计算和解微分方程.数值模拟方法是在分析方法对引力 N 体问题显得无能为力的时候,同时又在出现高速电子计算机的前提下,得以迅速发展.引力 N 体问题数值模拟用于星团、星系、星系团及整个宇宙的模拟研究,能较好地模拟一些天体系统,从而促进观测,推进恒星系统结构和演化研究的发展,因而倍受天体物理学家们的重视.计算机模拟毕竟是数值方法,它应该和分析方法结合起来,相辅相成,才能更好地推动引力 N 体问题,以至整个星系动力学的发展.

恩克方法(Encke method) 常微分方程组的经典数值方法之一.是德国恩克(Encke,J.F.)于19世纪初,在计算一个短周期彗星过程中建立的(此彗星后来被命名为恩克彗星).其特点是用直角坐标的摄动量作为基本变量,因其变化慢而可取较长的步长进行计算.

泰勒-斯蒂芬森方法(Tayler-Steffenson method) 常微分方程组的现代数值计算方法之一.在天体力学中常用到这种方法.其特点是将基本变量展开为自变量的泰勒级数,再根据情况截断,然后改写为数值方法的公式.实际应用表明,此法能节省计算时间.

解正则方程组的辛算法(symplectic integrator for the solution of canonical equations) 正则方程组数值计算的一种新方法.是在20世纪80年代由中国冯康首先提出的.第一篇论文是针对某些特殊正则方程组于1984年发表,后来又由他和国内外其他专家不断发展和完善,现在已得到公认.实际应用表明:此法最大优点是误差累计很慢,适合于长时间计算,而且天体力学中的大量动力演化课题都可以化为正则方程组,用此方法进行研究.

正规化变换(regularized transformation) 一种变换.指消除奇点的变换.在天体力学中常遇到的是碰撞奇点,而两体相碰撞的奇点是可以正规化的.这里以平面圆型限制性三体问题为例,常用到的正规化变换有:

1.列维-奇维塔变换.设(x,y)为旧坐标,t 为旧自变量;(X,Y)为新坐标,T 为新自变量.变换关系为:

$$x=\frac{1}{2}+X-Y,$$
$$y=2XY;$$
$$\mathrm{d}t=4(X+Y)\mathrm{d}T.$$

将此变换推广到三维空间,相应的变换称为KS变换,由库斯唐赫莫(Kustaanheimo,K.)和施蒂费尔(Stiefel,C.F.)于20世纪60年代提出.

2.蒂勒变换.旧变量仍为 x,y,t;新坐标用(E,F),新自变量用 T 表示.变换为:

$$x=\frac{1}{2}\cos E\,\mathrm{ch}F,$$
$$y=-\frac{1}{2}\sin E\,\mathrm{sh}F;$$
$$\mathrm{d}t=\frac{1}{4}(\mathrm{ch}F-\cos E)\mathrm{d}T.$$

列维-奇维塔变换(Levi-Civita transformation) 见“正规化变换”.

蒂勒变换(Thiele transformation) 见“正规化变换”.

天文动力学(astrodynamics) 亦称人造天体动力学.天体力学的一个新的分支.主要研究人类从地球上向空间发射的各种飞行器的运动,现在的人造天体基本上分为人造地球卫星、月球火箭和行星际飞行器三类.完整的人造地球卫星轨道分发射段、返航段和自由飞行段三部分.其中发射段火箭推力起主要作用;返航段指卫星再入大气层返回地面的回收飞行段,此时空气动力起主要作用.在天体力学范畴内,人造地球卫星运动理论主要研究自由飞行段的运动规律,为精密测定卫星轨道提供精确的动力学模型.月球火箭和行星际飞行器运动理论也是研究它们的自由飞行段的运动规律,为各种科学考察目的进行轨道设计.

人造天体动力学(artificial object dynamics) 即“天文动力学”.

人造地球卫星运动理论(theory of artificial satellite motion) 人造天体动力学的重要理论之一.它主要研究人造地球卫星在各种摄动因素影响下的运动规律.理论上沿用了经典天体力学中的摄动方法(即小参数方法)去求解卫星摄动运动方程.若地球是一个密度均匀的正球体,可把它看做质量集中于地心的一个质点,如果没有其他力的作用,人造地球卫星将围绕地心作椭圆运动.然而,它的运动却受到许多其他因素的影响,这些因素称为摄动.研究影响卫星运动的各种摄动因素及摄动运动方程的解法,是现代天体力学的一个重要研究课题.通过现代观测手段测定卫星的轨道还涉及数据处理与参数估计等问题,这又是天体力学与数理统计学交叉的一个课题.影响人造卫星运动的主要摄动因素有:地

球非球形摄动(即地球形状摄动);大气阻力摄动;太阳光压摄动;日、月引力摄动;还有因牛顿引力理论偏差产生的摄动,称为相对论改正或后牛顿摄动,现在已在测地卫星中研究并加以改正.

人造地球卫星的一阶运动理论(指包含到二阶长期摄动和一阶周期摄动的理论)通常采用分析方法进行研究,此时可将各种摄动因素分开处理.1959年,古在由秀采用平均要素法,首先提出一阶运动理论,后来被逐步完善.一阶分析解广泛用于精度要求不高的卫星定轨中.1962年,古在由秀首先创立二阶运动理论的分析方法,一般局限于地球非球形摄动.1970年,阿克斯内斯(Aksnes,K.)用包含部分一阶摄动的轨道作为中间轨道,并利用堀源-李变换,获得了简洁得多的表达式.由于二阶运动理论的公式繁杂,人们开始使用计算机进行公式的推导.现有的二阶分析解,采用的是简化了的模式,精度比数值方法要低.目前,国内外普遍采用数值方法(如亚当斯-科威耳方法)计算精密的卫星星历表,该方法用计算机实现较为方便,缺点是计算时间过长.为了解决分析方法精度不高而数值方法计算时间过长的问题,提出了半数值、半分析方法.短周期摄动用分析方法计算,长期与长周期摄动用数值方法计算,该方法对于计算分米级精度的卫星历表是合适的.

人造地球卫星轨道测定(orbit determination of artificial satellite) 天体力学中的一个研究课题.是主要研究如何由大量(如雷达、激光测距、多普勒测速、光电跟踪经纬仪等)观测资料求得卫星轨道根数及有关参数的最优估计的方法,包括初轨计算与轨道改进两部分.前者通常仅依据对卫星第一圈的观测资料,观测弧段短,且对轨道根数无先验估计;后者是依据对卫星若干天的观测资料,以初轨作粗估计,求得卫星轨道根数的精确估计.目前卫星定轨中普遍采用非线性的最小二乘估计.

在大量观测资料中常常包含部分异常值(指测量误差超过标准差若干倍的观测量),如何剔除大异常值,是数据预处理的重要课题.当数据中异常值的影响不可忽视时,应采用稳健参数估计的方法,代替古典最小二乘估计.

布劳威尔方法(Brouwer method) 人造地球卫星运动理论中的一种重要方法.是美国的布劳威尔(Brouwer,D.)于1959年提出,很快得到公认.特点是根据德洛内方法原理,用多次正则变换解决地球形状摄动问题.

发射轨道(ejective orbit) 天文动力学术语.指由地面发射飞行器到进入预定轨道的一段轨道.以火箭动力为主,地球引力和大气阻力也起重要作用.

同步轨道卫星(synchronous satellite) 亦称24小时卫星或固定卫星.天文动力学术语.定义为公转周期与地球自转周期相等的卫星.实际周期应为一个恒星日,等于23h56min4s或86164s.由此可算出同步轨道卫星到地心的平均距离为42164km,到地面的平均高度为35786km.但由于摄动作用,相对地面有微小飘移,需要经常用卫星上的制动装置予以改正.同步轨道卫星一般用作通信卫星、气象卫星、导航卫星等.

太阳同步卫星(solar synchronous satellite) 天文动力学术语.定义为轨道面法线方向与太阳方向的交角不变的卫星.由于地球公转,从地心到太阳的方向不断向东转动,每日约1°;又由于地球形状摄动,卫星轨道面也在转动,其角速度与轨道要素a(轨道半长径)、e(偏心率)、i(轨道面对赤道倾角)有关.只要在轨道设计中适当选择a,e,i的值,就可以使卫星轨道面的转动与太阳方向的变化一致,这就保证太阳同步卫星的实现.现在的太阳同步卫星主要是气象卫星.

过渡轨道(transfer orbit) 天文动力学术语.指从初始轨道转移到目标轨道的中间一段轨道.一般是偏心率很大的椭圆,在近地点处与初始轨道相切,在远地点处与目标轨道相切.发射近地点很高的卫星时(如同步通信卫星等),都是先进入近地点(离地面约200km)圆轨道,即初始轨道,然后在适当时刻点火加速,进入过渡轨道,到达远地点时再加速进入目标轨道.

地球非球形摄动(perturbation of nonspheric figure of the Earth) 非引力摄动之一.由于地球并不是一个正球体,其内部的密度分布不均匀而产生的摄动称为地球非球形摄动.目前,对地球的形状与密度分布主要采用三种近似模型:第一近似是假定地球为旋转对称体,相对自转轴分别对称,此时地球引力场的位函数与经度无关;第二近似是假定地球为有三个相互垂直对称面的对称体,此时地球引力场的位函数对纬度与经度分别对称;第三近似是假定地球是密度均匀的三轴椭球体.在第一近似下,位函数与时间无关,而在后两种近似下,由于地球自转,位函数与时间有关.地球非球形摄动使卫星不再沿一个固定的椭圆运动,它不仅使卫星轨道平面绕地球自转轴不断转动,同时使椭圆轨道在轨道平面内不停旋转.

大气阻力摄动(atmospherical drag perturbation) 非引力摄动之一.人造卫星在地球附近的空间飞行受到大气阻力产生的摄动称为大气阻力摄动.大气阻力摄动主要使卫星轨道的大小和形状不断变化,使卫星轨道越变越圆,且越变越小,最终使卫星在稠密大气中陨落.对于近地卫星来说,大气阻力是决定卫星寿命的主要因素.

太阳光压摄动(sunlight pressure perturbation) 非引力摄动之一.人造卫星被太阳照射时,受到太阳光压产生的摄动称为太阳光压摄动.由于地影的存在,卫星所受的太阳光压是间断的和不对称的,这将产生卫星运行周期与轨道偏心率的长周期变化.太阳光压摄动,对于面积质量比大的卫星,如气球卫星的运动,会起重要的作用.

日月引力摄动(lunisolar gravitational perturbation) 引力摄动之一.由太阳、月亮对人造地球卫星的引力产生的摄动称为日月引力摄动.对于近地卫星,日月引力摄动的量级较小;卫星越高,这种摄动就越大;到了地球同步卫星的高度,摄动就十分显著.日月引力摄动主要使卫星轨道产生长周期变化.

此外,影响人造地球卫星运动的摄动因素还有海潮摄动、固体潮摄动、大气潮摄动、月球扁率摄动、地球反射光压摄动、地球红外辐射摄动以及坐标系本身的运动所引起的附加摄动等,这些摄动因素通常都可忽略,只在特殊需要时才适当考虑.

坐标系摄动(perturbation of coordinate system) 天体力学术语.指非惯性坐标系中的惯性力改正.由于牛顿力学只能在惯性坐标系中成立,而实用的坐标系一般都是非惯性的(例如,用真赤道面作坐标面,在人造卫星运动中都是如此),必须改正坐标系摄动.

回收轨道(return orbit) 天文动力学术语.指从近地卫星安全返回地面的一段轨道.其间利用大气阻力和升力,再加上制动火箭推力,使飞行器缓慢减速逐渐安全降落地面.

轨道力学(orbital mechanics) 一种关于卫星的动力学.是以人造卫星观测为基础建立的卫星动力学.因地球引力场和大气摄动,很难建立精确的理论,而它们的大量参数就是从卫星观测资料定出的,故在20世纪80年代出现了以卫星观测来确定的力学模型,再由此模型建立卫星动力学的轨道力学.

航天飞机(space shuttle) 一种飞行器.能从地面发射进入卫星轨道,又能安全返回地面,并能多次重复使用的飞行器.

月球火箭运动理论(theory of the motion of lunar probes) 人造天体动力学的重要理论之一.是研究月球火箭在地、月引力及其他摄动力作用下的运动规律,为轨道设计提供依据的理论和方法.当前对月球火箭运动的研究主要用数值方法,而用分析方法进行定性研究,且为数值方法指明范围.分析方法主要采用所谓作用范围的概念.对于地月系统而言,月球的作用范围半径为66000km,火箭在此范围内飞行,可以认为只受月球引力的作用,它的轨道是以月心为焦点的圆锥曲线.反之,火箭在作用范围以外飞行,则只受地球的吸引,它的轨道是以地心为焦点的圆锥曲线.月球火箭的轨道分为:击中月球轨道、绕月飞行轨道、绕地-月飞行周期轨道和月球卫星轨道.

击中月球轨道(orbit for shooting at the Moon) 月球火箭轨道之一.为了击中月球,火箭的地心轨道可以是椭圆、抛物线和双曲线.但只有椭圆轨道既可以使火箭在到达远地点前从正面击中月球(上升轨道),也可以使火箭在过远地点后绕到月球背面去击中月球(下降轨道).其他两种轨道则只能从正面击中月球.在这三类轨道中,椭圆型轨道的稳定性最差,抛物线型轨道稳定性最高.谢多夫(Седов,Л.И.)的研究表明:从节省能量的观点来看,火箭进入轨道时的地心方向与火箭到达月球时的地心方向二者的交角愈大愈好,因此,火箭的飞行时间应在一天半、两天半或三天半左右.击中月球的轨道是一个典型的边值问题.

绕月飞行轨道(lunar orbit) 月球火箭轨道之一.火箭离开月球区域后能立即返回地球的邻近轨道称为绕月飞行轨道.这类轨道在月球附近,火箭对月心的速度要比月球抛物线速度(即逃逸速度)大一倍以上,因此,火箭相对于月球的运动总是双曲线型的.希勒对二维情形(火箭轨道在地月轨道面内)火箭的近月点分布进行了研究,发现在月球方向运动的前、后方各有一个不会有近月点的"禁区",前方的禁区比后方的大一倍.这说明考察月球的两侧要比考察月球的正、背面(尤其是正面)困难.切博塔廖夫(Чеботарёв,Г.А.)以平面圆型限制性三体问题为力学模型研究了绕月飞行的对称轨道,发现了一条对考察月球背面特别有意义的轨道.

绕地-月飞行周期轨道(periodic orbit revolved round the Earth-Moon system) 月球火箭轨道之一.通常以平面圆型限制性三体问题为力学模型来探讨绕地-月飞行的对称周期轨道.黄授书的研究表明:周期分别为

$$\frac{1}{2},\ \frac{2}{3},\ \frac{3}{4},\ \frac{2}{5},\ \cdots,\ \frac{6}{11}$$

个月的14种通约型轨道,它们的近月距小于8×10^4km,而近地距又在1.6×10^5km以内.这些近月距和近地距都不大的周期轨道对考察月球有实际意义.非对称周期轨道比对称周期轨道更稳定,但这类轨道比较难设计.

月球卫星轨道(orbit of lunar satellite) 月球火箭轨道之一.从地球上发射的火箭能否被月球俘获,变为月球卫星,这是一个俘获问题(参见"俘获理论").经研究证实,产生这种俘获的概率为零.因此,这种轨道的设计,只能从节约能量的角度,选择需要较小推力的轨道.月球卫星的运动将受到月球的形

状(非球形)摄动和日、地摄动等因素的影响.计算表明:月球周围 1×10^4km 的范围是卫星运动的稳定区,而逆行卫星轨道的稳定性比顺行轨道要好.

软着陆(soft landing) 天文动力学术语.是飞行器到达目标天体表面(包括返回地面)时,与天体表面的相对速度为零的情况.一般情况,虽然相对速度不是零,但飞行器能安全降落到天体表面,也认为是软着陆.

阿波罗月球探测(Apollo lunar exploration) 一种对月球的探测.这里指美国的庞大月球探测计划.于 1961 年开始进行预备工作,到 1972 年共发射了 17 次阿波罗号月球探测器.阿波罗 1,2,3 号是模型;4,5,6 号是不载人座舱试验;7,8,9,10 号为载人在地月间飞行作各种登月技术试验.1969 年 7 月 20 日发射的阿波罗 11 号,其登月舱第一次实现了载两名宇航员登上月球.之后的 12~17 号,都有两名宇航员参加登月,但其中 13 号发射失败,故阿波罗号共有六次 12 人登月成功.实现了在月面上安装激光反射器、人在月面行走、月面物质取样、月球地质地貌地震勘察,获得了月球矿产资源、近月空间环境等大量实测数据,所得资料至今尚未处理完.

月球激光测距(lunar laser ranging) 简称 LLR.一种天文观测技术.指由地面用激光测定月面某点的距离.自 1969 年 7 月 20 日阿波罗 11 号载人登月成功,并在月面安装激光反射器,就开始了月球激光测距工作.后来陆续在月面不同地点共安装 5 个激光反射器,其中有 1 个不成功.20 世纪 70 年代的测距精度为 8cm,到 20 世纪 90 年代已达 1cm,故激光测月技术已成为现代最精确的观测技术之一.用激光测月资料可精确测定地球自转参数、月历表偏差、月球天平动参数、引力理论参数等.

行星际飞行器运动理论(theory of the motion of interplanetary vehicle) 天文动力学研究领域之一.是主要研究行星际飞行器在日、地、目标行星引力及其他摄动力作用下的运动规律,依据科学考察的要求进行轨道设计的理论和方法.最终的轨道选择用数值方法计算,分析方法则利用作用范围的概念,主要进行定性研究,同时为数值方法提供近似轨道.行星际飞行器包括飞向和绕过行星的飞船、击中行星的火箭和行星的人造卫星等.利用作用范围的概念,可把行星际飞行器的运动分为三个阶段:从地球表面到离开地球作用范围之前,此时飞行器相对于地心的轨道应为双曲线;从离开地球作用范围之后到进入目标行星的作用范围之前为过渡阶段,主要研究飞行器的日心运动;以及进入目标行星的作用范围之后.行星际飞行器的运动主要是在轨道过渡阶段,这个阶段的轨道设计最为重要.最节省能量的过渡轨道是日心椭圆轨道,它在近日点和远日点上分别与相应的地球及目标行星的运行轨道相切,故又称双切轨道.这种轨道的飞行时间较长.如采用抛物线或双曲线轨道作为过渡轨道,可大大缩短飞行时间.事实上,目前发射的飞行器轨道大多属于双曲线类型.为了便于修正轨道和节省燃料,还设计了一种驻留轨道,它们是分别围绕地球和目标行星的卫星轨道.飞行器先发射到围绕地球的驻留轨道上,再加推力进入过渡轨道.在到达目标行星时也先在驻留轨道上运动,而后着陆.这样可以把暂时不用的燃料储存在驻留轨道上,返回时使用,从而大大减少能量的消耗.

水手号探测器(mariner probe) 一种行星际探测器.指美国发射的一组行星际探测器.主要用于探测火星、金星和水星.从 1962 年 7 月起到 1973 年 11 月为止,一共发射了 10 次,有 7 次成功.其中两次到金星附近,4 次到火星附近,一次在水星和金星轨道之间探测,发现了水星有球形磁场,其强度约为地球磁场强度的百分之一;还发现水星上有稀薄大气,含有氦、氖、氩等.

先驱者号探测器(pioneer probe) 一种行星际探测器.指美国发射的一组行星际探测器.任务是探测月球、水星、金星、木星和土星.自 1958 年到 1978 年的 20 年间,共发射成功 13 个.前几个探测内行星,发现地球磁层受太阳风压力成彗星形状.10 号探测器于 1972 年 7 月发射,12 月近木星后飞出太阳系.11 号探测器于 1973 年 4 月发射,1974 年 12 月飞近木星,1979 年 9 月飞近土星,发现两个土星新环,后飞出太阳系.1978 年还发射了两个先驱者号金星探测器,在金星表面软着陆.

旅行者号探测器(voyager probe) 一种行星际探测器.指美国发射的两个外行星探测器.旅行者 1 号于 1977 年 8 月发射,1979 年 3 月 5 日接近木星,发现木卫一的火山爆发、木星环、大红斑结构、新卫星木卫十四、十五、十六等.旅行者 2 号于 1977 年 7 月发射(比 1 号早),在 1979 年 7 月 9 日接近木星,发现木卫一上有七处火山、木卫二表面平坦并有冰层;1981 年 8 月 26 日飞近土星,发现六个新卫星;1986 年 1 月接近天王星,证实天王星环存在;1989 年 8 月飞近海王星,发现海王星环.两个探测器最后都飞出太阳系.为了可能被外星人得到,上面带有反映地球特点的一些照片,其中包括中国长城,还有十几种语言和乐曲的录音,其中有中国广东话和中国乐曲.

海盗号探测器(viking probe) 一种行星际探测器.指美国发射的 2 个火星探测器.任务是探测火星上的生命现象.海盗 1 号、2 号都在火星表面软着陆,分别于 1975 年 8 月 20 日和 9 月 29 日发射成功.结果未发现任何生命迹象,连有机分子都未

找到.

伽利略号探测器(Galilei probe) 一种行星际探测器. 指美国发射的木星探测器. 1989 年 10 月由亚特兰蒂斯号航天飞机携带升空,重 2550kg,1990 年 2 月飞掠金星,借助金星引力场第一次提高飞行速度并顺便考察了金星大气和云层. 1990 年 12 月和 1992 年 12 月两次折返地球附近,第二次和第三次加大飞行速度后才飞速奔向木星. 曾近距离探测了第 951 号小行星加斯普拉和第 243 号小行星艾达,1994 年 7 月接受临时指令考察了彗木碰撞. 1995 年 7 月 13 日,伽利略号所携带的木星大气探测器脱离飞船,于 12 月 7 日开始进入木星大气,在大气中约 150km 完成探测任务而陨落. 1995 年 7 月 27 日,伽利略号上的环木轨道飞行器启动马达,开始进入环绕木星的椭圆轨道. 它的主要任务是接收并储存木星大气探测器测定的木星大气的温度、压力、大气成分等以及它们随高度变化的数据,然后将数据发送回地球. 环绕木星 11 周,15 次飞掠伽利略卫星,对它们作近距离考察.

天体的形状和自转理论(theory of the figure and rotation of celestial body) 天体力学的一个分支. 研究天体在内外引力作用下自转平衡时的形状以及自转轴在空间和天体内部的变化规律,可分为形状理论和自转理论两部分. 形状理论以地球和其他行星为主要对象,采用转动的流体模型. 自转理论传统上采用刚体模型,现在也有采用刚体与流体或者刚体与弹性体相结合的模型. 研究得最多、最深入的是地球的自转,如岁差和章动、地极的移动和地球自转速率的变化等. 它是天文地球动力学的基本课题.

流体自转时的平衡理论(theory of equilibrium of rotating fluid) 研究天体形状的基本理论. 考虑天体的力学模型为自转的不可压缩流体,在平衡时可看做刚体. 对无辐射的天体(如行星)而言,流体内质点只受到自引力和自转离心力;对有辐射天体(如恒星),要加上辐射压力. 早在 18 世纪中期,法国克莱罗(Clairaut,A.-C.)和达朗贝尔(d'Alembert,J. le R.)就对行星情况用流体静力学建立起形状理论,初步解释了一些现象. 19 世纪后得到一系列的平衡形状. 有辐射情况是天体物理学的课题.

等位面(equipotential surface) 一种曲面. 天体形状理论中位函数等于常数时所确定的曲面称为等位面. 无辐射情况中,由流体自引力和自转离心力决定位函数,相应的等位面也是等密度面. 平衡形状的表面是一个等位面.

马克劳林椭球体(Maclaurin ellipsoid) 天体(看做均匀流体)自转时的一种平衡形状. 1742 年,马克劳林(Maclaurin,C.)首次证明:均匀流体达到自转平衡状态时,其形状可以是一个旋转椭球体,并且其赤道半径 a 必大于极半径 c,即是一个扁球体;并且证明

$$\Omega=\frac{\omega^2}{2\pi\sigma G}<0.22467\cdots=\Omega_0,$$

式中 ω 为自转角速度,σ 为流体的密度,G 为万有引力常数. 自转角速度越大,椭球体越扁. 但是,Ω 的最大极限为 Ω_0,后来,李亚普诺夫(Ляпунов,A. M.)又严格证明 $\Omega<\Omega_1=0.18711\cdots$ 时,马克劳林椭球体稳定,而 $\Omega_1<\Omega<\Omega_0$ 时为不稳定.

雅可比椭球体(Jacobi ellipsoid) 天体(看做均匀流体)自转时的一种平衡形状. 1834 年,雅可比(Jacobi,C. G. J.)证明:均匀流体自转达到平衡时可以是一个三轴椭球体,但是,其自转角速度 ω 必须是

$$\Omega<\Omega_1,\quad \Omega=\omega^2/2\pi\sigma G,$$

式中 ω 为自转角速度,σ 为流体的密度,G 为万有引力常数,$\Omega_1=0.18711\cdots$(参见"马克劳林椭球体"),这种三轴椭球体称为雅可比椭球体. 若 a,b 为椭球体赤道截面椭圆的半长径和半短径,c 为极半径,则有 $a>c$,$b>c$,说明是扁球体;当 $\Omega=\Omega_1$ 时,$a=b$,这时的雅可比椭球体成为马克劳林椭球体.

洛希极限(Roché limit) 天体力学的一个术语. 指流体卫星保持稳定平衡时与行星的最小距离. 法国天文学家洛希(Roché,E. A.)于 1849 年首先提出,故名. 卫星的形状与其主行星的引力作用有关. 若把卫星看做流体,由于主行星的吸引,卫星自转平衡时的形状为一个细长的、长轴方向指向行星的椭球体. 当卫星与行星的距离接近一个极限值时,这个椭球就会在引力作用下解体,这个距离的极限称为洛希极限,其值为

$$2.45539\left(\frac{\sigma'}{\sigma}\right)^{1/3}R,$$

式中 σ' 和 σ 分别为行星和卫星的密度,R 为行星的半径,系数 2.45539 是在假设卫星的质量与行星的质量的比值 $\mu=0$ 的情况下确定的,$\mu\neq0$,系数的值略有变化.

庞加莱极限(Poincaré limit) 天体力学术语. 是流体自转平衡时必须满足的一个条件:

$$\frac{\omega^2}{4\pi G\sigma}<1,$$

其中 ω 为流体自转角速度,G 是万有引力常数,σ 为流体密度.

木下宙方法(Kinoshita method) 现代用刚体模型建立的最精确的地球自转理论. 日本的木下宙首先于 1977 年提出. 特点是:用正则方程组表示地球自转运动方程,并用正则变换求解. 当时就得到公认,在 1979 年的国际天文学联合会大会上曾决定采

用，但因其他原因未能实现.后来发现方法中有些不够完善，1992年，进一步改进并提高精度，又受到国际上重视.

历书天文学(ephemeris astronomy)　天体力学的一个分支.其任务是研究如何用天体力学的理论和数值计算方法来编制各种天体的历表、预报天象以及建立天文常数系统.编算天文年历(参见“天文年历”)是历书天文学的一个重要内容.

贝塞尔日数(Besselian day number)　天文学术语.指天体平赤道球坐标同真赤道球坐标之间的转换关系中的五个系数.因是时间的函数，在天文年历中每日给一个值，称为贝塞尔日数.

贝塞尔岁首(beginning of Besselian year)　亦称贝塞尔假年岁首.天文学术语.天文上所用的贝塞尔年的年首.是贝塞尔(Bessel,F. W.)于19世纪初期提出的.为了天球坐标系之间的转换方便，贝塞尔用太阳平赤经增加360°定义为一年，还规定太阳平赤经加上光行差等于280°时作为贝塞尔假年岁首，它接近于回归年岁首.用年份后加符号“.0”表示.例如1990.0表示公历1990年的贝塞尔岁首，贝塞尔假年比回归年短0.001 48s.

地球时(terrestrial time)　简称TT.讨论近地天体运动所用的时间系统.1976年国际天文学联合会决定采用，命名为地球动力学时，代号TDT.并规定国际原子时TAI的1977年1月1日零时，为TDT的1977年1月1.000 3725日.其中0.000 3725d＝32.184s，是TAI起点1958年1月1日零时历书时ET与世界时之差.这使TDT和ET衔接起来，并有

$$TDT=TAI+32.184s.$$

20世纪90年代后，因考虑到TDT同动力学没有联系，故简化为地球时.

地球动力学时(terrestrial dynamical time)　见“地球时”.

质心动力学时(barycentric dynamical time or temps dynamique barycentrique)　讨论行星系统运动所用的时间系统.1976年，国际天文学联合会决定采用，命名为(太阳系)质心动力学时，代号TDB.由于引力场和运动速度不同，TDB与地球时TT的差非常复杂.为了简化起见，规定TDB的起点和秒长与TT相同，故二者只有周期差.在一般问题中可忽略，现在用的太阳系精确历表，自变量就是TDB.近年TDT改为TT后，TDB也相应简化为TB.

力学时(dynamical time)　一种时间系统.指地球动力学时TDT和质心动力学时TDB.如不加说明，力学时指TDT，即TT.

食(eclipse)　天文学术语.一个天体被另一个天体的影子全部或局部遮掩的天文现象称为食.当月球从太阳和地球之间经过，三者近似在一条直线时，月影扫过地区的观测者可看到日食.只有月球在其轨道交点之一附近和太阳相合(朔)时才发生日食.月球(或地球)沿着与太阳相反的方向投射出圆锥形影锥，月球(或地球)与太阳的外切线构成本影锥，内切线构成半影锥.处于月球本影锥内的人完全看不到太阳，这时发生日全食；处于半影锥内的人，只能看到太阳的一部分，这种现象称日偏食.由于月球到地心的距离是变化的，当月球远离地球时，本影锥不能达到地面，其延长部分(伪本影)经过的地区只能看到太阳的边缘，便发生日环食.当地球在月球和太阳之间，三者近似成一直线时，月球没入地影锥，便发生月食.只有月球靠近其轨道交点之一并和太阳相冲(望)时，才能发生月食.如果月球完全没入地球本影锥内，则称月全食；如果部分月面被本影遮掩，则为月偏食；若月球只进入半影锥，则称为半影月食.日食只能在地面上一狭窄地带内被观测到，月食则在半个地球范围内可同时观测到.就整个地球而言，见到日食次数多于月食.就一个地区来说，月食多于日食.全世界每年见食次数平均4次.最少2次，都是日食；最多7次，日食5次，月食2次，或日食4次，月食3次.木星的四个大卫星(伽利略卫星)绕木星运行进入木星影锥内的现象称木卫食.

掩星(occultation)　天文学术语.指月球在绕地球运行过程中遮住遥远的恒星或行星遮掩其卫星的现象.最常发生的是月掩恒星，通常称月掩星.月掩太阳则称为日食.恒星或行星的掩始发生在月球东边缘，复现于西边缘.一次掩星只能在地球上一定的区域被看到.在该区域之外，由于视差效应使恒星或行星不与月面接触.观测月掩星可以测定月球坐标，也被用来测定观测站的地心坐标.

凌日(transit)　天文学术语.指行星或彗星通过太阳表面的现象.一个卫星通过母行星表面的现象则称卫凌行星.只有水星和金星有凌日现象，水星和金星的轨道与黄道的倾角分别为7°和3.4°，所以只有当水星(或金星)和地球都很接近水星(或金星)的轨道的一个交点，并且必须是下合(参见“行星动态”)时才能发生凌日.地球经过水星轨道交点发生在11月10日或5月8日前后，所以水星凌日只能发生在这两个日期附近.同理，金星凌日只能发生在6月7日或12月9日附近.每100年内平均发生水星凌日13次.20世纪80年代和20世纪90年代发生水星凌日的日期为1986年11月13日、1993年11月6日和1999年11月15日.下两次水星凌日将发生在2003年5月7日和2006年11月8日.上一次金星凌日发生在1882年12月6日，下两次将发生在2004年6月8日和2012年6月6日.木星

的四个伽利略卫星除木卫四偶然从木星圆面旁边通过外，每逢其中任一颗与木星下合时都发生木卫凌日.

晨昏蒙影(twilight) 一种天文现象. 指傍晚从日没到天黑和早晨从天黑到日出这一段时间天空发亮的现象. 在日出前的称晨光，在日没后的称昏影. 这种特殊的亮光是地球高层大气散射日光所引起. 晨昏蒙影持续的时间与观测者的地理纬度、经度和海拔高度有关，又与季节有关，它也与当地自然条件特别是气候条件有关. 习惯上分为三级：太阳在地平线下 6°时称民用晨光始或民用昏影终，一般进行户外作业时无需人工照明；太阳在地平线下 12°时称航海晨光始或航海昏影终，这时能看见地平线和较亮的星；太阳在地平线下 18°时为天文晨光始或天文昏影终，阳光在天空中完全消失. 在赤道，晨昏蒙影持续时间较短，随着地理纬度的增加，蒙影时间增长，在高纬度地区夏至时民用昏影刚结束，民用晨光随即开始，整夜出现晨昏蒙影，称为白夜. 纬度越高，发生"白夜"的日子越多.

白夜(white night) 见"晨昏蒙影".

沙罗周(saros) 天文学术语. 指日月食发生的一种周期. 沙罗是巴比伦文的"saros"，意为"重复"或"恢复". 古巴比伦人发现每次交食经过 18 年 11 天(6585.3 天)或 18 年 10 天(若这期间有 5 个闰年)后重又出现. 这段时间包含 223 个朔望月(或 242 个交点月或 19 个太阳交点年)，即经过一个沙罗周，太阳、月球与黄白交点差不多，又回到原来的相对位置，因此，前一周期的日月食又重新出现. 每一沙罗周内平均有 71 次交食，其中日食 43 次，月食 28 次.

天文年历(astronomical ephemeris) 天文学用的一种历表. 按年度出版的反映天体运动规律的历表称为天文年历. 目前，天文年历的主要内容有：太阳、行星、月球的精确位置表；上千颗主要恒星的位置表；各种天象的预报，如日食、月食、月掩星、行星动态、日月出没和晨昏蒙影等；一些重要的天文常数和用于天体各种坐标和时间之间换算的必要数表. 有的天文年历有几颗较大的小行星位置表及木星四颗大卫星的历表. 1679 年，法国出版的《关于时间和天体运动的知识》是世界上最早的天文年历. 1767 年英国、1776 年德国开始逐年出版天文年历，美国(1855)、苏联(1941)、中国(1966)相继独立编算出版天文年历.

DE200/LE200 **数值历表**(DE200/LE200 numerical ephemeris) 天文学用的一种历表. 是现在多数国家天文年历中计算日月行星位置所依据的基本历表. 美国航天局在 20 世纪 60 年代根据行星际航行导航的需要，向喷气推进实验室(JPL)提出了研究更精确的太阳系历表的任务. JPL 将此取名为动力学历表(dynamicalephemeris)，简称 DE 系列. 到 1982 年完成时，相应序列为 DE200 及对应的月球历表(lunar ephemeris)LE200，故总称为 DE200/LE200. 此历表是用数值方法算出日月大行星在 1820～2060 年间的精确位置. 运动方程是根据参数化后牛顿(PPN)方法列出的，其中考虑了日月九大行星的相互引力、地月形状摄动、月球天平动以及五个质量最大的小行星摄动. 1984 年起，正式被各国天文年历采用. 但在应用中感到精度仍不够，近年来又完成了新的系列 DE245，DE303，DE343 等.

VSOP82**分析历表**(VSOP 82 analytical ephemeris) 天文上所用的一种历表. 是法国和比利时合作的大行星分析历表. 在其展开式中，角变量和行星轨道要素都表示成时间的幂级数，系数为数值，故为半分析历表. 1983 年完成后，同 ELP2000 月历表一起成为一些国家 1984 年后天文年历中日月大行星位置计算的根据. VSOP82 的适用年限是 1800～2050 年. 另有 VSOP85 精度低些，但可用到公元前 2000 年，便于对古代观测资料(如古代天象记录等)的研究.

航海历(nautical almanac) 专用历书之一. 根据《天文年历》编算，专供航海人员观测天体以决定船舰地理坐标的一种专用历书. 又称航海天文历. 最初为《天文年历》的组成部分，后来为了改进航海技术而单独出版，主要内容有：太阳、月亮以及金星、火星、木星、土星四大行星的格林尼治时角和赤纬(一年内每隔 1 小时列出一值)，少数恒星的格林尼治时角和约 60 颗天体赤纬(一年内每隔 3 天列出一值)；173 颗辅助航海星的每月视位置，日出、日没、月出、月没、黄昏、黎明的时刻，天文现象日记、日月食的预报等. 天体位置角度的精度为 0.1′，相当于 0.1 海里的精度. 第一部专用的《航海天文历》于 1914 年由英国出版，以后 10 年中各主要国家也相继出版，中国于 1954 年编算出版《航海天文历》.

航空历(air almanac) 专用历书之一. 根据《天文年历》编算的供航空人员观测天体以决定飞机位置的一种专用历书. 又称航空天文历. 基本内容有：太阳、月亮以及金星、火星、木星和土星四大行星的格林尼治时角和赤纬(一年内每隔 1 小时列出一值)，少数恒星格林尼治时角和较亮恒星的视位置. 中国于 1955 年编算出版《航空天文历》.

大行星运动理论(theory of major planets) 一种有关天体运动的理论. 是研究大行星运动的方法和理论. 大行星运动是人们最早关注并着手研究的天体运动. 托勒密(Ptolemy)的地心说和哥白尼(Kopernik，M.)的日心说都曾是阐述行星运动的极有代表性的学说. 1609～1619 年，开普勒(Kepler，

J.)在分析总结了第谷(Tycho Brahe)的大量火星观测资料后提出了行星运动三定律(参见“开普勒定律”),将行星的运动用严格的数学语言和公式进行描述.但开普勒定律对行星的运动只有现象的归纳而无力学本质的解释.自从牛顿(Newton,I.)发现了万有引力定律以后,在此基础上建立了行星运动方程以及开展了如何积分这种方程的研究,这时行星的运动理论才算正式建立.行星主要受太阳的引力作用,同时也受其他行星的引力摄动,研究如何积分行星受摄运动方程是大行星运动理论的核心,目前分析方法和数值方法都被采用.大行星运动理论是编制行星历表的基础,天文年历登载了九大行星的历表.目前,中国采用美国用数值方法计算的DE200/LE200 历表,法国等西欧国家用 VSOP 分析历表.

水星近日点反常进动(Mercury perihelion unusual advance) 一种天文现象.指按牛顿引力定律计算的水星近日点进动值与观测值的差异.1859年,法国天文学家勒威耶(Le Verrier,U.J.J.)观测测定水星近日点进动速率,比根据牛顿引力定律计算出的进动速率快 38″/世纪.他认为可能是未知的水内行星的引力所致,但一直未发现水内行星.1882年,美国天文学家纽科姆(Newcomb,S.)求出反常进动速率为 43″/世纪,后曾陆续提出一些理论来解释反常进动均未成功.1915 年,爱因斯坦(Einstein,A.)发表了广义相对论,根据这一理论,水星近日点进动与由牛顿引力定律计算出的差值为 43.03″/世纪,与观测值十分接近.这成为广义相对论的最有力的经典验证之一.

拉普拉斯不变平面(Laplace invariable plane) 天体力学的一个概念.指太阳系总动量矩(包括公转和自转)的垂直平面,由拉普拉斯(Laplace,P.-S.)在 18 世纪首先提出而得名.如只考虑质点之间的万有引力,根据质点组动力学的动量矩定理可知,质点动量矩之和为常量,即大小和方向都不变(称为质点组的总动量矩),因而与总动量矩垂直的平面也是不变平面.如把太阳系看做质点组,拉普拉斯不变平面差不多与木星轨道面平行.

天然卫星的运动(motion of natural satellite) 天体力学研究课题之一.天然卫星是太阳系中围绕行星运动的小天体,它们的运动轨道比太阳系其他天体更复杂,公转方向有顺行的,有逆行的;轨道偏心率有很小的,也有很大的.摄动理论的分析方法和数值方法都可以用来研究天然卫星的运动,但要根据具体对象作不同的选择.卫星除受主行星的引力作用外,主要的摄动来自太阳和卫星之间的相互吸引.卫星绕行星运动时如同行星绕太阳的运动,会发生轨道共振,如木卫六和木卫七的平均角速度 n_6 和 n_7 有以下关系:

$$3n_6 - 4n_7 \approx 0;$$

有时还有双共振,如木卫一、木卫二、木卫三,三颗卫星的平均角速度之间有以下关系:

$$n_1 - 3n_2 + 2n_3 \approx 0.$$

很多卫星距主行星很近,它们对主行星的潮汐作用也不可忽略.潮汐作用改变了主行星的形状和引力场,又反过来影响卫星的运动,使得有一类卫星的轨道变大,而另一类卫星的轨道变小,直至陨落.近年还发现木卫和土卫中有几乎共轨的卫星,更增加了运动的复杂性.

月球运动理论(theory of motion of the Moon) 一种天然卫星运动理论.这里指研究月球运动的理论和方法.月球是地球惟一的天然卫星,也是离地球最近的一个天体.影响月球运动的因素很多,因此,月球的运动很复杂,通常分两步讨论:第一步,将太阳、地球和月球看做质点,研究月球在地球和太阳引力作用下的运动,称为主问题;第二步,再考虑地球、月球为非正球体的形状摄动、行星的引力摄动、太阳相对地月质心的非椭圆轨道的影响以及潮汐摩擦等.通常人们所说的月球运动理论是指研究主问题的各种理论和方法.1687 年,牛顿(Newton,I.)提出万有引力定律时,首先用它研究了月球的运动.自18 世纪以后相继提出的各种月球运动理论多达几十种,其中较好的有十几种,以布朗(Brown,E.)改进的希尔理论较为完善,他编制的《月球运动表》自1923 年后被天文年历所采用.1960 年后直接用他的理论公式计算.现代月球历表有两种,美国、中国等多数国家采用美国的 LE200 数值历表,西欧一些国家采用 ALE 2000 分析历表.月球运动的分析解中,包括各种长期变化项和周期项,多达几百项.布朗的月球运动表中就有 300 多种不同的周期项,称为月行差.月球轨道变化有两项主要的长期项:一是月球轨道升交点西退,周期为 18.6135a;另一项是月球近地点有周期为 8.8475a 的东移.

月球轨道升交点西退(retrograde motion of ascending node of the Moon orbit) 描述月球运动的指标之一.指月球运动的一项长期变化项.月球绕地球运动时,由于太阳的引力摄动,轨道升交点不断西退,西退周期为 18.6135a.交点的这一运动直接影响日、月食的计算.中国古代早已发现了这个周期,称为一章.

月球近地点东进(forward motion of the Moon perigee) 描述月球运动的指标之一.指月球运动的主要长期变化项.在太阳的引力摄动下,月球的近地点有周期为 8.8475a 的向东移动,即每 8.8475 年近地点转一圈.近地点的这一运动引起月地距离的变化,因而对潮汐产生一定的影响.

月球天平动(libration of the Moon)　天文学术语. 指的是地面观测者所观测到的月球可见面上下左右小幅度的摆动. 由于天平动效应,使观测者能看到月面的 59%,其中 18%时多时少. 因月球运动和自转轴指向而造成的天平动,又称光学天平动或几何天平动;因月球实际摆动形成的天平动则称物理天平动. 光学天平动分为三种:

1. 经度天平动. 月球公转轨道为一椭圆,公转速度不均匀,而月球自转速度却是均匀的,所以月面每月(近点月)在偏离中心位置大约 7°54′而东西摆动.

2. 纬度天平动. 月球自转轴不与其轨道面垂直,因此在其公转过程中,月球自转轴轮流朝向地球,形成月面每月(交点月)偏离中心位置最大约 6°51′而上下俯仰.

3. 视差天平动. 由于视差的原因,地面观测者在月出、月没时,各能多看到月面东边缘外侧和西边缘外侧约 1°的部分区域. 月球以很小的振荡来回摆动,称物理天平动,约 2″.

二均差(variation)　天体力学术语. 指由于太阳引力摄动引起的月球黄经变化中的一项周期变化,周期为半个朔望月,振幅为 39′29.9″.

出差(evection)　天体力学术语. 指月球黄经变化中最大的周期项,是太阳的引力摄动引起的. 该项的周期为 1.80747d,振幅为 1°16′26.4″.

周年差(annual equation)　天体力学术语. 为月球运动中的主要周期摄动项之一. 此项周期为一近点年,月球黄经中此项的振幅为 11′8.9″.

月角差(parallactic inequality)　天体力学术语. 指月球运动中的主要周期摄动项之一. 此项周期为一朔望月,月球黄经中此项的振幅为 2′4.8″.

月球运动加速现象(acceleration phenomenon of lunar motion)　一种天文现象. 指月球运动中有重要历史作用的一种观测现象. 早在 18 世纪中期,天文学家们就发现由观测得到的月球运动总是比理论计算的运动快,称为月球运动的加速现象. 到 19 世纪末,纽科姆(Newcomb,S.)经 20 余年的研究,肯定了月球加速与时间成正比,并发现内行星也有同样结果. 他正式提出这不是月球和内行星理论不准,而是以地球自转为基础的时间变化在不断减慢,也就是地球自转在不断减慢. 20 世纪初得到公认,成为天文学中的一项重要发现. 原子钟出现后,完全证实了该结论. 后来发现,由于地月系的潮汐演化,月球运动反而有减速现象,也已被观测肯定.

ELP2000 **月历表**(ELP2000 lunar ephemeris)　天文学用的一种历表. 指法国编制的月球分析历表,与 VSOP82 行星分析历表配套,也是半分析的,其精度比希尔-布朗方法的历表提高了两个数量级,1983 年完成.

多体问题(many-body problem)　亦称 N 体问题. 天体力学的基本问题之一. 它研究以质点表示的天体系统中 N 个天体在万有引力作用下的运动规律. 多体问题是一种特殊的质点系动力学,是一般力学的专门分支. 每个质点的质量、初始位置和初始速度可不加任何限制,或因所研究的力学系统不同而有所限制. $N=2$ 的二体问题是最简单的多体问题,它的运动方程早在牛顿时代已基本解决,即两个天体的轨道或一个天体相对另一天体的轨道都是圆锥曲线. 而三体问题,三百多年来历经很多著名的科学家的艰苦努力仍未解决. 就连针对具体天体系统提出的限制性三体问题仍未完全解决. 由于电子计算机的发展,对于 N 大于 3 的 N 体问题,现在主要使用数值方法,例如,外行星四百年的坐标,就是用计算机算出来的. 一般地,要综合运用分析方法、定性方法研究 N 体问题. 由于计算技术的迅速发展,近年已用数值方法研究 $N>10000$ 的多体问题,用于对星团或星系理论模型的研究.

N 体问题(N body problem)　即“多体问题”.

经典积分(classical integrals)　天体力学术语. 指 N 体问题(N 为大于 1 的任意正整数)中的运动方程组的 10 个首次积分:6 个质心运动积分;3 个动量矩积分;1 个能量积分. 这些积分在 18 世纪就已求出,称为经典积分. 它们是四个守恒定律(质心运动守恒、动量守恒、动量矩守恒和总能量守恒)的体现.

对合系统(involution system)　天体力学术语. 多体问题或其他动力系统的首次积分组成的一个特殊系统,其中任意两个组合的泊松括号恒等于零. 按照李氏(Lie,M. S.)定理,若有 k 个首次积分组成对合系统,则用它们可以把原动力系统降低 $2k$ 阶.

位力定理(Virial theorem)　亦称维里定理. 多体问题的经典定理. 该定理断言:稳定多体系统应满足关系

$$2T+U=T+E=0,$$

其中 T,U,E 分别为多体系统的总动能、总势能、总机械能. 若多体还有自转和磁性,则式中 T 应加上自转动能,另外再加上总磁能.

中心构形(central configuration)　多体问题的一种特解. 在此解中,N 体运动要永远保持一定的对称形状,如正多边形解和共线解等(参见“三体问题定性研究”).

雅可比坐标系(Jacobi coordinate system)　一种特殊坐标系. 指多体问题研究中常用的一种相对坐标系. 一般取最大质量为第一体. 讨论第二体运动时取第一体为坐标原点;讨论第三体运动时取第一、

第二体的质心为坐标原点;如此继续下去,最后讨论第 N 体运动时,取前 $N-1$ 体的质心为原点. 坐标轴是平行于某惯性系的三个坐标轴.

限制性问题(restricted problem) 一种特殊的多体问题. 设多体系统中有 N 个大天体和 k 个小天体. 大天体对系统内所有天体的引力都要讨论,但小天体对大天体的引力可以忽略,只讨论小天体之间的引力,故又称为 $N+k$ 体问题. 此课题是 20 世纪 80 年代才提出的,适用于研究小行星系和卫星系的运动.

三体问题(three-body problem) 天体力学研究的重要课题之一. 指对三个天体(质点)在万有引力作用下运动的研究,而它们的质量、初始位置和初始速度是任意的. 这是三百多年来仍未解决的一大难题. 研究三体问题的方法有分析方法、定性方法和数值方法. 分析方法中的级数解法是建立带有小参数形式的三体问题运动方程,把天体坐标或轨道要素表示为时间的函数,展成级数形式的近似分析表达式,用以讨论天体的运动规律. 由于收敛的限制,只能用于较短的时间间隔. 两百年前,拉格朗日(Lagrange,J.-L.)利用 10 个初积分将三体的 18 阶微分方程降为 6 阶,这就提醒人们降阶法是求解方程的一个思路,但是至今没有进展. 此外,寻找三体问题的特解或周期解的问题,也没有获得突破性进展. 有人转向特殊三体问题的研究,即所谓限制性三体问题. 定性方法是用微分方程的定性理论研究三体问题的全局特性. 数值方法是直接算出天体在某时刻的具体位置,有利于实际应用,但难以研究天体运动的变化规律. 实际上,往往综合应用几种手段来研究三体问题会更为有效.

三体问题积分(integral of three-body problem) 探索一般三体问题的新积分的研究. 一般三体问题的运动方程是 18 阶的微分方程组,需 18 个初积分才能求解,但至今只知道 10 个初积分,因此,寻找新的积分成为求解三体问题的重要途径,然而却得到种种悲观的结论. 1887 年,布伦斯(Bruns,H.)证明:如用坐标和速度作基本变量,则三体问题不存在新的代数积分. 1889 年,庞加莱(Poincaré,(J.-)H.)又证明:如用轨道要素的组合作变量,则新的单值解析积分也不存在. 1898 年,班勒卫(Painlevé,P.)证明:表示为速度分量的代数函数形式的新积分也不存在. 这就提示人们必须从新的途径寻找新积分. 直到 1966 年,希腊研究组用数值探索得出结论:平面圆型限制性三体问题可能存在新积分.

三体问题幂级数解(power series solution of three-body problem) 三体问题的一种解. 指宋德曼(Sundman,K. F.)于 1913 年提出的一般三体问题的解,此解可以展开为自变量的幂级数,并证明在时间趋于无穷时,级数收敛. 此定理在数学上得到很高评价,但因收敛太慢而在天文学中不能应用.

三体问题的定性研究(qualitative study of three-body problem) 天体力学定性理论研究领域之一. 指研究有关三体问题可积性和级数解存在性及收敛性等定性内容(参见"三体问题"). 现着重介绍下列应用领域:

1. 三体问题的定形解. 在圆型限制性三体问题中有 5 个平动点,在惯性坐标系中相应为 5 个圆轨道特解,其中三个解与两个有限体保持直线,两个解与两个有限体保持等边三角形. 在一般三体问题中,类似的特解仍然存在. 由拉格朗日(Lagrange,J.-L.)于 1772 年提出. 5 个特解在惯性坐标系中都是圆锥曲线轨道,若为椭圆时,三体的运动周期相同. 由于这些特解使三体保持一定形状,故称为定形解,并分别称为等边三角形解和共线解.

2. 奇点和正规化. 三体问题的奇点只有一种,即碰撞奇点. 两体碰撞奇点可用自变量变换消除,这种变换称为正规化变换;三体同时碰撞对应的奇点至今尚未找到能消除的变换. 现已证明,三体同时碰撞前的短暂时间内,三体必然组成等边三角形或共线,正好与定形解相对应. 在接近碰撞附近,用数值方法计算也会遇到困难,而正规化变换可以解决. 另外,即使不产生碰撞,两体紧密接近时轨道会有较大变化,原来相对运动是椭圆轨道的可能变成双曲线;相反情况也会发生,这就是俘获或交换.

3. 时间趋于无穷时的运动特性. 设 r_{12},r_{23},r_{31} 为三体相互间距离,在时间 $t\to+\infty$ 时,按距离变化情况,可分为 7 类运动:双曲型($r_{ij}=O(t)$);双曲-抛物型($r_{12}=O(t)$,$r_{23}=O(t)$,$r_{31}=O(t^{2/3})$);双曲-椭圆型($r_{12}=O(t)$,$r_{23}=O(t)$,$r_{31}<C$ (C 为某确定正数));抛物型($r_{ij}=O(t^{2/3})$);抛物-椭圆型($r_{12}=O(t^{2/3})$,$r_{23}=O(t^{2/3})$,$r_{31}<C$);椭圆(有限)型($r_{ij}<C$);振动型($r_{12}<C$, r_{23},r_{31} 无界,但不趋于无穷). 再按三体情况可细分为 16 类,称为沙士分类. 在时间 $t\to-\infty$,有同样的 16 类,因此,若讨论三体从 $t\to-\infty$ 到 $+\infty$ 时的运动特性,共可分为 $16^2=256$(类),归纳了三体所有可能的运动情况,称为三体问题的终结运动. 但到现在为止,各种运动类型的条件并未全部讨论清楚,只是根据三体的总能量常数 $h<0$,$h=0$, 或 $h>0$ 时,分别判别了哪些类型可能存在. 在椭圆型中,周期轨道和拟周期轨道情况研究得最多.

4. 运动区域问题. 20 世纪 70 年代以后,由于找到了一般三体问题的希尔曲面,用它可以讨论相应的运动区域范围及运动流形的拓扑结构. 在三体总能量 h 为某些特殊值时,相应的希尔曲面出现奇点,相应的流形拓扑结构会出现变化.

一些特殊的三体问题近年来也研究得很多，如毕达哥拉斯问题、西特尼科夫问题等，可参阅相应条目.

三体问题降阶(reduction of three-body problem) 使三体问题运动方程降阶的理论研究. 利用10个经典积分，一般情况下只能把18阶的三体运动方程降低到8阶. 但由于三体问题的特点，可降低到6阶. 拉格朗日(Lagrange, J.-L.)于1776年首先求出，后来雅可比(Jacobi, C. G. J.)、庞加莱(Poincaré, (J.-)H.)等人又用另外的方法，也把三体问题的运动方程降到6阶.

三体问题的拉格朗日特解(Lagrange special solution of three-body problem) 亦称三体问题的定形解. 见“三体问题的定性研究”.

限制性三体问题(restricted three-body problem) 三体问题的特殊情况. 即一个天体质量与其他两个天体质量相比，小到可以忽略的三体问题. 小行星、彗星和人造地球卫星在太阳和一个大行星吸引下运动的问题可作为限制性三体问题来研究. 此时，小质量天体对两个大质量天体的引力作用是不必考虑的，两个大质量天体的运动就是二体问题的运动，其运动轨道为圆锥曲线. 因此，限制性三体问题中，只讨论小天体的运动. 如果两个大质量天体的运动轨道为圆，则此问题称为圆型限制性三体问题. 同样，如果两个大质量天体的运动轨道为椭圆(或双曲线，或抛物线)，则称为椭圆型(或双曲线，或抛物线)限制性三体问题. 在月球运动、脱罗央小行星群的运动的研究中，常按限制性三体问题来处理.

平面圆型限制性三体问题(plane circular restricted three-body problem) 限制性三体问题中比较简单、研究得最多的一类. 它假定两个主星体(有限质量)在相互作用下绕其质量中心作圆周运动，只研究运动在主星体轨道平面内的第三个小天体(无限小质量)在两个主星体的万有引力作用下的运动规律. 当取主星体连线为 x 轴时，则相应于旋转坐标系的小天体运动方程为

$$V^2=x^2+y^2+\frac{2m_1}{r_1}+\frac{2m_2}{r_2}-C,$$

其中 m_1，m_2 分别为主星体的质量，r_1，r_2 分别为小天体到主星体的距离，(x,y)，V 分别为小天体的坐标和速度，C 为积分常数，这就是雅可比积分. 当小天体的速度为零时，就有零速度方程

$$x^2+y^2+\frac{2m_1}{r_1}+\frac{2m_2}{r_2}=C.$$

积分常数 C 决定于小天体的初始速度和初始位置，积分常数 C 的取值不同，表示小天体的不同运动范围形成 L_1，L_2，L_3，L_4 和 L_5 5个平动点，这是小天体运动方程的5个特解. L_4，L_5 和两个主星体构成两个等边三角形，太阳系中脱罗央小行星群就是一个著名的例子. 平面圆型限制性三体问题在月球火箭运动和双星演化的研究中得到进一步的运用.

梯塞朗判别式(Tisserand criterion) 证认彗星的一种判别式. 因大多数彗星受木星的摄动，轨道要素变化很大，公转一圈或数圈再回到太阳附近后，很难判定是否为同一颗彗星. 法国天文学家梯塞朗(Tisserand, F.)于19世纪末提出判别式

$$\frac{1}{a}+2\sqrt{a(1-e^2)}\cos i=C,$$

其中 a，e，i 分别为彗星轨道的半长径、偏心率和对黄道的倾角. 如两次观测到的彗星的 a，e，i 代入上式后得到的 C 值相同，则为同一彗星.

毕达哥拉斯(三体)问题(Pythagorean (three-body) problem) 一种特殊的三体问题. 三体在初始时刻静止在直角三角形 ABC 顶点上，其中 C 点处为直角；三边 BC，CA，AB 的长度分别为3，4，5个单位；A，B，C 处三体的质量也是3，4，5个单位. 然后三体在万有引力作用下运动，研究这种三体在时间 t 趋于无穷时的运动状态，就称为毕达哥拉斯问题. 研究此问题的目的是数值探索的检验. 20世纪70年代，由美国和西欧两个研究组用同一类型计算机，同样的数值计算方法，算到很长时间后，两体成为双星，另一体远离而去. 只是两个研究组得到的三体成员不同：一组是 B，C 体成双星，A 体远离；另一组是 A，C 体成双星，B 体远离. 可能是误差传播造成的结果.

等腰三体问题(equilateral three-body problem) 限制性三体问题的一种特殊情况. 设两体 A，B 质量相等，相互围绕质心(即二体连线中点 O)作椭圆运动，再设另一体 C，质量为无穷小(只受 A，B 两体吸引，而可略去 C 对 A，B 的引力)，初始条件使得它在通过 O 点垂直于 AB 连线的平面内运动，因而永远有 $AC=BC$，使三体成为等腰三角形. 它可作为一种恒星系统的简化模型，也是三体问题定性研究的一个特殊模型. 它的一种特例又称为西特尼科夫问题.

西特尼科夫问题(Sitinikov problem) 等腰三体问题的特例. 质量为无穷小体 C 永远在通过 A，B 两体(质量相等)质心(连线 AB 的中点 O)，并垂直于 A，B 两体运动平面的直线上. A，B 两体绕中点 O 作椭圆轨道运动，故又是椭圆限制性三体问题的一种特例. 由于它只有一个自由度，又是可积系统，近年来常用于数值探索的检验.

双不动中心问题(two fixed-center problem) 一种特殊的限制性三体问题. 在限制性三体问题中，如果两个主星体的相对运动很微小，它们对空间某一参考系来讲，可以看做是固定不动的，小天体在这

两个主星体的引力作用下运动.这个问题最早由欧拉(Euler,L.)提出,故也称欧拉问题.双不动中心问题是一个可积系统,因此可作为一种中间轨道(参见“中间轨道”),用来研究人造卫星在地球引力场中的运动.

哥本哈根问题(problem of Copenhagen)　研究周期轨道有系统结果的一个著名问题.20世纪30年代,哥本哈根天文台斯特龙根(Strömgren,B.G.D.)提出了一个计划,用数值方法找出平面圆型限制性三体问题(两有限体质量相等)的几种周期轨道族:一种是围绕两个有限体附近;另一种同时围绕两个有限体;第三种是围绕5个平动点附近,并讨论这些周期轨道族之间的联系和演化关系.1936年,在哥本哈根召开的国际会议上肯定了这个计划.后来的20多年间,这些周期轨道族一般都找到了.研究这类工作的群体又称为哥本哈根学派.

碰撞问题(collision problem)　天体力学中的一类研究课题.讨论 P 个天体在时刻 t 碰撞于一点时的运动性质的理论.当时间趋于 t 时,天体距离 r 趋于零,运动方程产生奇点,能否通过一定的变换消除这一奇点,碰撞以后天体如何运动,在碰撞时刻附近轨道的渐近表现如何,以及虽不发生碰撞但几个天体互相紧密接近,这时轨道性质如何等,均属碰撞问题的内容.二体碰撞问题中通过一定的变换可消除奇点(称为正规化变换),多体碰撞问题则复杂得多,还有许多问题尚未解决,至今尚未找到相应的正规化变换.

相对论天体力学(relativistic celestial mechanics)　天体力学研究的新领域.即用爱因斯坦的广义相对论或现代其他新引力理论为基础来研究天体运动的学科,是天体力学和引力物理学相结合的一个新领域.长期以来,天体力学都是以牛顿力学为基础.虽然勒威耶(Le Verrier,U.J.J.)于1859年已发现水星近日点进动速率的观测值比用牛顿力学得到的计算值要大38″/世纪,但他怀疑是存在水内行星所引起的.1896年,纽科姆(Newcomb,S.)把此差值订正为43.03″/世纪,并对牛顿万有引力定律提出怀疑.1916年,爱因斯坦(Einstein,A.)提出的广义相对论解释了这个差值:理论值比牛顿力学结果增加43.03″/世纪,当时观测值订正的差值为43.11″/世纪.这表明广义相对论完美地解释了水星进日点进动现象,误差只有0.08″/世纪,与观测误差一致.但限于当时的观测精度,天体运动的一般课题仍以牛顿力学为基础,个别精度特别高的课题才以广义相对论的近似结果作改正,由此建立了后牛顿天体力学,作为后来建立的相对论天体物理学的一个领域.20世纪60年代以来,由于观测新技术的发展,精度大幅度提高,用牛顿力学建立的天体运动理论与观测结果的偏差日益显著,用广义相对论或更新的引力理论来建立天体运动理论开始提到议事日程.同时,测出的太阳扁率较大,会使水星近日点进动增加4″/世纪,这样,对广义相对论也产生了怀疑.于是在十几年间,相继出现了上百种与广义相对论有区别的引力理论.但到20世纪80年代,新理论纷纷被否定,只剩下个别的理论能与广义相对论并存.另外,更多的太阳形状观测表明:太阳的扁率更小些,对水星近日点进动的贡献只有0.3″/世纪左右,引起的矛盾不大.广义相对论进一步得到公认,新的天体运动理论应该以广义相对论为基础.

第一次提出“相对论天体力学”这个学科名词的人是勃龙别格(Брумберг,В.А.),他在1972年出版的《相对论天体力学》一书中,较系统地讲述了此学科的基础和主要课题.但因用俄语出版,国际上影响不大,到20世纪80年代才逐步引起重视.美国和西欧各自建立起以广义相对论为基础的行星和月球运动理论,作为新出版的天文年历的根据(1984年起),但都精确到后牛顿近似,现仍在继续改进.在西方学者要求下,勃龙别格编写了《相对论天体力学》新的英文版(已于1991年在英国出版),他被选为第19届国际天文学联合会天体力学专业委员会的主席(1985—1988).作为一个学科,相对论天体力学尚在建立过程中,但已形成了自身较系统的基础理论和系列课题,并已取得一批成果.基础理论分为:

1.相对论质点组动力学,即多体问题.现已建立起(1981年)参数化后牛顿(简称PPN)多体问题运动方程及相应的参考系,高阶后牛顿理论只有二体问题.

2.相对论延伸体动力学.即讨论各种形状和内部结构的天体在相对论框架内的运动.对二体问题讨论较详细,现正建立完整的后牛顿延伸体多体问题(由法、德和中国科学家合作).

基本课题分两大类:一是相对论太阳系动力学,讨论行星、月球、人造卫星等太阳系天体在相对论框架内的运动.目前已建立美国的行星和月球数值历表(称为DE系列)和西欧行星和月球分析历表(VSOP系列和ALE系列),但还不符合精密观测(如激光测月等)要求.二是相对论恒星系统动力学,研究双星、星团、星系及宇宙动力学.其中双星,特别是脉冲双星动力学研究较多,是检验相对论的新目标.宇宙动力学也很活跃,因为动力学与其他物理过程常结合在一起,所以这个领域与相对论天体物理学交叉较多.

后牛顿天体力学(post-Newtonian celestial mechanics)　简称PN天体力学.研究天体运动的一门学科.是用广义相对论的一阶近似理论研究天体运动的学科.严格的广义相对论研究天体运动遇到困

难，故在场方程中按某小量（与光速平方成反比）进行展开，不含此小量的理论就是牛顿力学．用含此小量的一阶理论（简称 PN1）所建立的天体运动理论称后牛顿（或一阶后牛顿）天体力学．由此可类推二阶后牛顿（PN2）或更多阶后牛顿理论．

参数化后牛顿方法（parametric post-Newtonian formalism） 简称 PPN 方法．检验不同引力理论的一种常用方法．20 世纪 60 年代后，出现了很多不同于爱因斯坦广义相对论的引力理论．为了检验，由威尔（Will，C. M.）、诺德威特（Nordtvedt，K. J.）等人建立起 PPN 方法，即在后牛顿理论中加入 10 个参数，不同理论的参数值不全相同，再由观测或实验测定这些参数，以判定符合哪种理论．

天体物理学

天体物理学（astrophysics） 天文学的一个分支．它应用物理学的理论和方法研究天体的物理状态、化学组成、内部结构和演化规律．可分为实测天体物理学（测量太阳、恒星、行星、星云的电磁辐射的能量分布和偏振，以及研究它们的光谱）和理论天体物理学（用理论物理学的原理和方法，根据实测天体物理学提供的观测资料来研究天体的物理性质和物理过程）．按所用的物理理论，又可细分为高能天体物理、相对论天体物理、等离子体天体物理等．天体物理学还可按研究对象分为太阳物理学（包括与太阳有关的全部现象和日地关系）、太阳系物理学（包括行星大气和行星内部的性质，彗星、流星的形成，黄道光的粒子的化学和物理状态，行星际等离子体及磁场，地球和木星的磁层等）、恒星物理学（包括恒星大气的物态、内部结构、元素的合成，产能机制，恒星的演化，以及星云和星际物质的物态等）、恒星天文学（研究大量恒星的集体特性，求出恒星的空间分布和运动规律、恒星系统的结构和动力学、起源和演化）、星系天文学（研究星系的结构、运动、起源和演化，星系、星系群、星系团的空间分布、相互作用以及星系际物质的物态等）、天体演化学、天体化学等．天体物理学应用现代天体物理的理论、技术和方法研究天体，宇宙天体的各种极端物理状态为物理规律的验证提供天然实验室，天体物理学的新发现又激励和促进物理学的进步和发展．

实测天体物理学（observational astrophysics） 天体物理学的一个分支学科．它应用物理学的理论和实验技术研究天体物理学中的基本观测仪器的原理和结构以及观测结果的处理方法以获取天体信息．除极少数太阳系天体外，因人们无法直接接触天体，故对天体的研究主要依靠收集和分析来自天体的电磁辐射．由于波长不同的电磁辐射，有时具有截然不同的物理性质，因此发展起不同的天文仪器和观测技术．如对可见光采用人眼、照相底片、光电倍增管、电荷耦合器件作辐射接收器，对 γ 射线则采用计数管．此外，还要研究星际介质和地球大气对天体辐射的消光作用以及消除它们影响的理论和方法．根据天体辐射的强度、频谱、偏振、能谱等可以测定天体的温度、压力、密度、视向速度、电磁性质以及化学组成等，从而为理论天体物理学提供研究资料．

天体光度学（astronomical photometry） 实测天体物理学的一个分支学科．它利用测光技术测量和研究天体的亮度（来自天体的有限波段范围内的辐射流），确定天体的光度、光谱型和距离．测光是研究天体物理性质的重要方法．天体光度测量，根据所用的技术分为目视测光、照相测光、光电测光等，是目前应用最广的测光方法．用电荷耦合器件测光灵敏度高．对有一定视面的天体，如星系、星云、太阳、月球和行星，测光又分为研究天体视面亮度分布和测量整个天体视面的累积星等．目前，现代最大的望远镜用目视可看到 18 等星，用照相底片能拍摄到 23 等星，用电荷耦合器件可拍摄到 25 等或更暗的星，照相测光的误差约为 0.04 星等，光电测光精度可达 0.01 星等．

（恒星的）光度（luminosity (of stars)） 表征恒星发光能力强弱的量．恒星表面每秒钟辐射的总能量称为光度．它是恒星本身的真正发光能力——真亮度．恒星光度常用绝对星等 M 表示，可利用

$$M = m + 5 + 5\lg\pi - A(r)$$

求出，式中恒星视星等 m 和恒星视差 π 及星际消光改正 $A(r)$ 可由实际测量得出．有时用太阳光度作单位，这时恒星光度 L 可由公式

$$L = \frac{2.512^{4.7-M}}{L_{\odot}}$$

求出，其中 $L_{\odot}$ 为太阳光度，M 为恒星绝对星等．恒星的光度彼此相差悬殊，有的恒星的光度比太阳大几十万倍（超巨星），例如天蝎座 ι'（$M=-8.4$ 等）、剑鱼座 S（$M=-8.9$ 等）；有的恒星的光度只有太阳光度的五万分之一（矮星），例如沃尔夫 359（$M=+16.68$ 等）．

（天体的）亮度（brightness (of celestial body)） 表征天体辐射流大小的量．观测者所接收到的天体辐射流的大小，即天体辐射流在辐射接收器处的照度称为亮度．天体亮度 E 常用星等 m 表示．星等 m 与天体亮度 E 的对数成正比，即 $m=-2.5\lg E$，E 以零等星的照度 E_0 为单位，据现代测量结果，1 标准国际烛光在离观测者 1m 时，星等为 -13.98 等，求得零等星的照度相当于 2.54×10^{-6}lx．

星等（magnitude） 天文学术语．是天文学中表示天体亮度的量，通常用 m 表示．古希腊天文学家

喜帕恰斯(Hipparchus,(R.))首先将全天人眼可见的恒星按感觉的亮度分成6等,最亮的星定为1等星,次亮的称2等星,依次类推.肉眼看见的最暗的星为6等星.后来普森(Pogson,N.R.)确定,星等差 m_1-m_2 和恒星亮度比的对数成比例,即

$$m_1 - m_2 = - 2.5\lg \frac{I_2}{I_1} .$$

星等尺度的零点由规定某颗星的星等数值来确定.各种辐射接收器对不同波段的辐射的响应灵敏度各不相同,因此有不同的星等系统,如目视星等、照相星等、光电星等.对有视面的天体,其总亮度所相应的星等称为累积星等.照相星等是用对蓝光(波长为4270Å)敏感的天文底片测定的星等.目视星等介于5.5～6.5等之间的A0型星,其照相星等和其目视星等相等.用正色照相底片加黄色滤光片(波长5900Å)所测定的星等,因与目视星等相近称为仿视星等.光电星等是用光电倍增管测定的星等.由天体在整个电磁波段内的总辐射而确定的星等称热星等.

照相星等(photographic magnitude)　见“星等”.

目视星等(visual magnitude)　亦称视星等.一种表征天体亮度的量.指用人眼测定的天体亮度(星等),以 m_v 表示.人眼视网膜对可见光的黄色光(有效波长为5500Å)敏感.美国哈佛大学天文台规定,以小熊座λ星的目视星等 $m_v=+6.55$ 等作为目视星等标准,据此可以确定目视星等的零点.例如,太阳目视星等为−26.74等,天狼星为−1.6等.目视星等为1等的星,在地面的照度约等于

$$8.3\times10^{-9}\text{lx}.$$

仿视星等(photovisual magnitude)　一种表征天体亮度的量.指用正色照相底片加黄色滤光片测定的星等.滤光片滤去蓝光和紫光而让黄、绿光通过,结果得到与目视星等相近的星等,故名仿视星等,以 m_{pv} 表示,它的有效波长为5900Å.

热星等(bolometric magnitude)　一种表征天体亮度的量.指由天体在整个电磁波谱内的总辐射而确定的星等,以 m_b 表示.定义为

$$m_b=-2.5\lg E+c',$$

式中 E 为天体亮度,

$$E = \int_0^{\infty} E_\lambda \mathrm{d}\lambda,$$

c' 为零点常数.将天体想象地置于10秒差距时的热星等,称为绝对热星等 M_b,

$$M_b=-2.5\lg L+c_b,$$

式中 L 为天体的总辐射通量,c_b 为零点常数.M_b 与 m_b 有如下关系:

$$M_b=m_b+5-5\lg r-A_b(r),$$

r 为天体距离,$A_b(r)$ 为热星等的星际吸光改正值.用温差电偶、辐射测热计等辐射探测器可测定天体的热星等.为普遍求得大量天体热星等,引入物理量热改正(BC).热改正定义为

$$\mathrm{BC}=m_b-m_v=M_h-M_v .$$

求得热改正后,可将目视星等换算成热星等.

绝对星等(absolute magnitude)　一种表征天体光度大小的量.是为了比较天体的真正发光能力,把天体假想地置于10秒差距处所测得的目视星等.绝对星等 M 与恒星视星等 m、恒星距离 r 的关系为

$$M=m+5-5\lg r-A.$$

实测出恒星的距离 r 和视星等 m、改正星际消光 A 后,可以求出绝对星等 M.也可由恒星的光谱或恒星的颜色(色指数)测定恒星的绝对星等(参见“分光视差”),太阳的绝对星等为+4.7等.

色指数(colour index)　天文学术语.指天体在任意两个波段内的星等差.早期把照相星等 m_{ph} 和目视星等 m_v 的差称为色指数,用 c 表示,$c=m_{ph}-m_v$.规定星等在5.5～6.5等之间的A0型恒星其色指数为零,蓝色恒星色指数为负值,红色恒星色指数为正值.现在通用的色指数是宽带UBV测光系统中的B−V和U−B.色指数是一表征恒星颜色的量,恒星颜色是由恒星色温度 T_c 决定的.如果测出色指数 c,可由下式求得恒星的色温度 T_c:

$$T_c = \frac{7200}{c + 0.64}.$$

采用UBV三色测光系统,天体的色温度 T_c 与色指数B−V则有如下关系:

$$T_c=\frac{7900}{(\mathrm{B-V})+0.72}.$$

由于星际物质对星光的选择散射,对短波散射大,对长波散射小,因此恒星颜色要红,称为星际红化.色指数要经过星际红化改正求出真色指数,用来确定恒星的色温度.经过星际红化改正后的真色指数与光谱型是很近似于一一对应的,因此常用真色指数来代替光谱型绘制赫罗图.未经过星际红化改正的色指数与相应光谱型的平均色指数之差称为色余.

星际红化(interstellar reddening)　见“色指数”.

色余(colour excess)　见“色指数”.

热改正(bolometric correction)　天文学术语.指热星等 m_b 与目视星等 m_v 之差,常以BC表示.定义为

$$\begin{aligned}\mathrm{BC} &= m_b - m_v \\ &= - 2.5\lg \left(\int_0^{\infty} E_\lambda \mathrm{d}\lambda\right) - m_v + c_1 \\ &= 2.5\lg \left(\int_0^{\infty} E_\lambda v_\lambda \mathrm{d}\lambda \Big/ \int_0^{\infty} E_\lambda \mathrm{d}\lambda\right) + c_2,\end{aligned}$$

式中 E_λ,v_λ 分别是天体的辐射流量和分光灵敏度,c_1,c_2 为零点常数.近代实测结果为:

$$c_1=-11.51\pm0.03\text{等},$$

$$c_2=0.958\pm0.01\ \text{等}.$$

各类恒星的热改正值可查表求出.

多色测光(multicolor photometry) 一种测光系统.指用有色滤光片或干涉滤光片和光电倍增管,或用照相方法测量天体在不同波带内的辐射能,来研究恒星物理性质的一种测光系统.选择不同波长的滤光片和辐射探测器,可组成具有各种平均波长 λ_m 和通带半宽 $\Delta\lambda$ 的多色测光系统.由于通带半宽 $\Delta\lambda$ 的不同,可分为宽带多色测光($\Delta\lambda>300$ Å)、窄带多色测光($\Delta\lambda<100$ Å)、中带多色测光(300 Å $>\Delta\lambda>$ 100 Å).宽带多色测光由于通带半宽大,获得天体的能量较多,可用来测定暗星连续光谱的能量分布,求得它们的色温度和进行光谱分类;窄带多色测光得到的信息较多,能测量相邻很近的两光谱区的强度比,可代替分光光度测量中的某些工作,常用来测量吸收线;中带多色测光介于二者之间.国际上常用的多色测光系统有 UBV 三色测光系统及 uvby 四色测光系统等.通过多色测光,能得到恒星的色指数、色温度、有效温度,甚至可求出绝对星等、恒星大气中金属元素含量和恒星表面重力加速度等物理参量.

UBV **测光系统**(UBV photometric system) 用光电方法测量恒星光度的一种测光系统.即应用有色滤光片和光电倍增管,测量恒星在紫外波带(平均波长 3500 Å,通带半宽 600 Å,以 U 表示)、蓝光波带(平均波长 4300 Å,通带半宽 950 Å,以 B 表示)和黄光波带(平均波长 5500 Å,通带半宽 1400 Å,以 V 表示)的光度.20 世纪 50 年代,首先由美国约翰逊(Johnson, H. J.)和摩根(Morgen, W. W.)在麦克唐纳天文台使用.这种测光系统已作为国际测光标准系统.UBV 测光系统的星等值常用 V 星等和色指数 B−V 及 U−B 来表示.

四色测光系统(four-colour photometry system) 亦称 uvby 测光系统.一种中波带多色测光系统.由丹麦天文学家斯特龙根(Strömgren, B. G. D.)于 20 世纪 50 年代提出.这一系统使用四种滤光片配以光电倍增管测量恒星的光度.u 指紫外,平均波长为 3500 Å,通带半宽 300 Å;v 指紫光,平均波长为 4110 Å,通带半宽 190 Å;b 为蓝光,平均波长为 4670 Å,通带半宽 180 Å;y 为黄光,平均波长为 5470 Å,通带半宽为 230 Å.现在,四色测光系统观测的星数仅次于 UBV 测光系统,在四色测光系统中,测光获得的结果由下面三个量给出:

$$\mathrm{b-y},$$

$$c_1=(\mathrm{u-v})-(\mathrm{v-b}),$$

$$m_1=(\mathrm{v-b})-(\mathrm{b-y}).$$

色指数 b−y 与 UBV 测光系统中的色指数 B−V 相当,c_1 是巴耳末跳跃的一种量度,m_1 则与金属线的多寡有关.由于四色测光的通带要比 UBV 系统窄得多,在作消光改正时,几乎与恒星的色指数无关,因此它的归算要比 UBV 系统简单得多.分析四色测光系统,能够得到恒星的许多基本物理参量,例如绝对星等、温度和表面重力加速度等.这种测光系统还可进行恒星的二维或三维分类,确定星际红化、粗略测算晚型星的元素丰度以及获得其他许多物理量,它可以得到比 UBV 三色测光系统更多的信息.

uvby **测光系统**(uvby photometric system) 即"四色测光系统".

北极星序(north polar sequence) 天文学术语.指北天极附近的恒星光度星表.由勒维特(Leavitt, H. S.)于 1917 年刊布,有 96 颗星,照相星等从 2.55 等(北极星)到 20.10 等.1922 年,被国际天文学联合会作为国际的照相星等标准.有人又根据美国黑尔天文台(威尔森山)所测定的仿视星等换算了北极星序里的恒星的仿视星等,从 2.08 等到 17.43 等,也被国际天文学联合会作为国际仿视星等标准.它的不足之处是:在南半球完全观测不到,而对北半球的中、低纬度,北极星序的恒星天顶距太大,受大气消光影响大,不利于精确测光.

测光标准星(photometric standard stars) 天体物理学术语.指作为测光标准的一类星.在一定测光系统中,其精确亮度和色指数为已知,可用来作为测定其他恒星的亮度和色指数的恒星称为测光标准星.北极星序中 96 颗星曾作为测光标准星.20 世纪 50 年代建立的 UBV 测光系统中给出 108 颗测光标准星,后来又给出一些疏散星团中的测光标准星.1965 年,斯特龙根(Strömgen, B. G. D.)发表了 uvby 四色测光系统中 1217 颗星的测光结果,可看做是测光标准.1973 年,已编制出版了 7600 颗恒星的测光星表.

天体照相学(astrophotography) 天文学的一个分支学科.是照相方法在天文学上的应用.照相是收集、存贮和传递信息的一种手段.其优点是:照片能长期保存,具有文献性;照片提供了客观的可存贮的星象,不会因观测者主观印象而改变,具有客观性;长时间露光可观测到肉眼看不到的暗星以及天体的红外和紫外辐射;能用来记录天体的光谱及特征形态.天体照相学的主要内容有:天体测量(测定恒星位置、自行和三角视差);光度测量(依据照片上星象的大小和密度测定视星等);照相巡天观测(用广视场照相机观测提供全天空的图象);观测天体外貌特征(用专用望远镜拍摄太阳表面大气,用空间探

测手段拍摄天体的X射线、紫外、红外波段的图象等).

天体光谱学(astrospectroscopy) 实测天体物理学的一个分支学科.它研究和利用分光技术,拍摄和分析天体的光谱,并与实验得到的标准谱线比较,来确定天体的物质结构、大气的物态、化学组成以及天体的运动状况.天体光谱是研究天体的主要信息,是天体物理学研究的基本资料.根据天体光谱可确定天体大气中的元素及其丰度.由谱线的多普勒位移和宽度,可以测定天体的视向速度、自转、脉动、谱线生成区及分光双星运动情况.根据谱线的强度和能量分布,可以确定天体大气的物理状态、电磁过程等.

(恒星的)光谱型(spectral type (of star)) 天体物理学术语.指按恒星光谱特征进行的分类系统.恒星的辐射光谱是连续光谱叠加夫琅禾费吸收线,温度极高的恒星和特殊星的光谱还伴随有发射线.根据谱线种类和它们的强弱程度可将恒星光谱分类.通常采用美国哈佛大学的光谱分类法,将恒星分为O,B,A,F,G,K,M七大类型,并将少数恒星分为S型和R,N型.对每一光谱型又细分为10个次型,以数字0,1,2,…,9标记,如太阳光谱是G2型.为了说明恒星的特点,还在光谱型前标记小写字母,如c表示超巨星,d表示矮星,g表示巨星,e表示光谱有发射线等.如太阳为dG2,表示太阳为G2型矮星.各光谱型的主要特征(强吸收线及吸收带)及相应的温度如下表所示,表明光谱型序列

光谱型	表面温度/K	强吸收线和吸收带
O	50000	电离氦,三次电离硅
B0	25000	中性氦,二次电离硅
A0	11000	中性氢,电离硅
F0	76000	中性氢,电离钙
G0	6000	电离钙,中性铁
K0	5100	中性金属
M0	3600	中性金属,TiO
M5	3000	中性金属,TiO
R,N	3000	中性金属,CN,C_2
S	3000	中性金属,ZrO

```
                      ┌─S
O—B—A—F—G—K—M
                      └─R—N
```

为形成吸收线的恒星大气的温度序列.

天体分光光度测量(astronomical spectrophotometry) 天体物理学研究的内容之一.指对天体某一波长的单色辐射的测量.可分为单色分光光度测量和多色分光光度测量.前者为测量很窄波段上的光度,用于测定光谱线的能量分布;后者则对不同波段进行光度测量,用于研究连续光谱的能量分布.

色温度(colour temperature) 亦称分光光度温度.表示恒星连续光谱在某一波段的能量分布的一个量,即与天体在某一波段内连续谱能量随波长的分布相近的绝对黑体所相应的温度.当天体和绝对黑体的辐射特性相近时,色温度相当精确地表示天体发出辐射的实际温度.在UBV测光系统中,测出色指数B－V,则由下式可求出天体在4300～5500Å波段内的色温度T_c,即

$$T_c=\frac{7900}{(B-V)+0.72}.$$

或用分光光度测量方法,将某天体在某一波段的连续谱与一已知色温度的天体的连续谱进行比较归算后,也可求得天体的色温度.1919年,波茨坦天文台维尔兴(Wilsing,J.)等首次用分光光度测量方法定出199颗恒星的色温度.

分光光度温度(spectrophotometry temperature) 即"色温度".

谱线轮廓(spectral line profile) 描述光谱谱线内能量分布的一种曲线.即谱线内某一波长的辐射强度(I_λ)与连续光谱在同一波长的辐射强度(I_λ^0)之比(称为谱线剩余强度r_λ)随波长变化的曲线.实际应用中,有时用谱线深度R_λ.R_λ的定义为$R_\lambda=1-r_\lambda$.谱线最中心部分R_λ最大,剩余强度r_λ最小,对应于吸收线最暗的部分.

仪器轮廓(instrumental profile) 天体物理学术语.指由于观测仪器的狭缝宽度和衍射现象引起的谱线轮廓.假定谱线是无限窄的,观测到的谱线总是有一定宽度的谱线轮廓,这种轮廓即仪器轮廓.必须从观测到的谱线轮廓中消去仪器轮廓的影响,才得到真谱线轮廓.设$I(x)$表示离谱线中心为x处的真轮廓,$A(\lambda-x)$为仪器轮廓,$O(\lambda)$为线心波长λ处的观测轮廓,它们之间由积分方程

$$O(\lambda)=\int_{-\infty}^{+\infty}A(\lambda-x)I(x)\mathrm{d}x$$

相联系.由于谱线真轮廓具有各种形式,并无确定的变化规律,而仪器轮廓又没有一定的函数形式表示,它随具体的仪器不同而不同,因此作仪器轮廓改正比较困难.改正方法有佛克特(Voigt)函数近似法、迭代法、图算法和傅里叶变换重建过程等.

等值宽度(equivalent width or equivalent breadth) 天体物理学术语.指谱线内被吸收的总能量的一种度量.定义为

$$w_\lambda=\int_0^\infty(1-r_\lambda)\mathrm{d}\lambda=\int_0^\infty R_\lambda\mathrm{d}\lambda$$

或

$$w_\nu = \int_0^\infty (1 - r_\nu) \mathrm{d}\nu,$$

w_λ 称为等值宽度，r_λ 称为谱线剩余强度，R_λ 为谱线深度.等值宽度代表了谱线所吸收的全部能量，故又称谱线总吸收.因为高度为一个单位，宽度为 w_λ 的矩形面积和总吸收的面积相等，这时矩形宽度在数值上和总吸收相等，故总吸收又名等值宽度.

生长曲线(curve of growth)　定量分析恒星吸收光谱时使用的一种曲线.天体光谱中某一吸收线的等值宽度(表征谱线吸收总能量的一种量)与恒星大气内吸收辐射而产生该吸收线的原子数有关.以等值宽度(w_λ/λ)的对数作纵坐标，处于某一能态和某一体积内的某元素参与吸收的原子数(f)的对数为横坐标，所得出的表示二者关系的曲线称为生长曲线.分别以

$$\lg\left(\frac{w_\lambda}{\lambda}\frac{c}{v_0}\right), \quad \lg\left(\frac{\sqrt{\pi}e^2}{m_e c \Delta\nu_D} N_i f\right)$$

为纵坐标和横坐标，画图所得的表示二者关系的曲线称为理论生长曲线.其中 c 为光速，v_0 为原子热运动平均速度，m_e 为电子质量，e 为电子电荷，$\Delta\nu_D$ 称为多普勒宽度，N_i 表示 $1\mathrm{cm}^3$ 内处在 i 能态的原子数.将生长曲线和理论生长曲线比较，可以求出原子热运动平均速度 v_0(从而算出热运动温度 T)、N_i(从而求出化学组成)以及恒星大气的其他物理量.

分光视差(spectroscopic parallax)　天体物理学术语.指利用恒星光谱特性测定的恒星视差.1902年，赫茨普龙(Hertzsprung，E.)首先发现电离锶谱线(Sr Ⅱ 4078)的强度和绝对星等有关.例如，同一光谱型的矮星(太阳)和超巨星(猎户座ζ)的光谱谱线几乎完全一致，惟有 Sr Ⅱ 的两根谱线在猎户座ζ的光谱内比在太阳光谱内强6000倍，中性钙谱线(Ca Ⅰ 4277)在矮星光谱内非常强，在超巨星光谱内非常弱.1914年，美国威尔逊山天文台亚当斯(Adams，W. S.)等建立了测定恒星视差的分光方法：对每一光谱次型，选取一些已知其三角视差的恒星，以绝对星等 M 为纵坐标，以对绝对星等敏感的一些谱线对强度之比为横坐标绘图，据此，由恒星某些敏感谱线的强度可求出恒星的绝对星等 M.进一步求出恒星的视差.对视差小于0.075″的恒星，分光视差法比三角视差法准确.对于视差大于0.075″的恒星，则用三角视差法测定比较准确.分光视差法只能用于F，G，K，M型恒星，B型和A型恒星的光谱中金属谱线微弱或完全不出现，但可由氢线的强度和轮廓来定出A型和B型的绝对星等.

秒差距(parsec)　一种天文学单位.指天体周年视差为1″时对应的距离.是天文学中测量天体距离的单位之一，以pc表示.Parsec由parallax(视差)和second(秒)两字的头几个字母组合而成，原意为“视差为(角)秒(的距离)”，中译秒差距.利用星的视差来测量天体距离的设想是德国天文学家贝塞尔(Bessel，F. W.)等人于1840年提出来的，秒差距这个名词据说是英国天文学家特纳(Turner，H. H.)提出的.在1922年的国际天文学联合会第一届大会上得到承认.

1 秒差距 $=3.085\,678\times10^{13}$ 千米
$=206\,264.8$ 天文单位
$=3.261\,631$ 光年.

表示遥远的星系、类星体的距离时常用千秒差距(kpc)为单位.最近的恒星半人马座比邻星与地球的距离为1.3秒差距，仙女座星系与地球的距离为680千秒差距.

光年(light year)　一种天文学单位.指光在真空中运行一年所行经的距离.是天文学中计量天体距离的一种单位，1888年首次使用.

1 光年 $=9.4605\times10^{12}$ 千米
$=63\,240$ 天文单位
$=0.3066$ 秒差距.

最近的恒星半人马座与邻星距离为4.22光年，仙女座星系距离为220万光年.有时以光传播的时间表示太阳系天体距地球的距离，如月球距地球的平均距离为1.3 s，太阳距地球的平均距离为8.3min，火星最靠近地球时距地球的距离为3.1min，木星最靠近地球时距地球的距离为33min，冥王星最靠近地球时距地球的距离为5.3h.

天体视向速度测量(determination of astronomical radial velocity)　天体物理学研究课题之一.它根据天体光谱谱线的位移确定天体沿视线方向运动的速度，是研究天体运动的一种有效方法.可直接测定出天体运动的视向速度分量，其准确度不受天体距离的影响.根据多普勒原理，光源和观测者的相对运动会影响观测者所接收到的辐射波长，使辐射谱线位移.位移 $\Delta\lambda$ 的大小和视向速度 v 的关系为

$$\Delta\lambda = \frac{v}{c}\lambda,$$

式中 c 为光速，$\Delta\lambda=\lambda'-\lambda$，$\lambda$ 为光源不动时的光波波长，λ' 为光源在视线方向相对运动时所观测到的波长.用照相方法拍摄天体光谱，精密测定运动光谱谱线波长 λ'，再与不动光源对应的谱线波长 λ 进行比较，即可得出 $\Delta\lambda$，从而进一步测出视向速度.

(天体的)视向速度(radial velocity (of celestial body))　一种天体物理量.指的是天体相对观测者视线方向的运动速度分量.即天体趋近或远离观测者的速度.视向速度使天体光谱谱线发生位移.根据多普勒原理，光源和观测者的相对运动会影响观测

者所接收到的辐射波长，即谱线位移量 $\Delta\lambda=\lambda'-\lambda$（$\lambda$ 表示光源不动时的光波波长，λ' 为光源在视线方向相对运动时所观测到的波长）的大小和视向速度 v 成正比，即

$$\Delta\lambda=\frac{v}{c}\lambda,$$

式中 c 为光速. 通常当天体远离观测者时，v 取正值，$\Delta\lambda>0$，由视向速度引起的谱线向红端位移；当天体趋近观测者时，v 为负值，$\Delta\lambda<0$，谱线向紫端位移. 利用天体分光学测量天体的谱线位移可以测定天体的视向速度.

天体偏振测量(astronomical polarimetry) 天体物理学研究课题之一. 内容为研究确定天体电磁辐射的偏振特性的仪器和测量方法等. 1811 年，法国科学家阿拉果(Arago，D. F.)发现月球反射的太阳光是线偏振光，此后陆续发现行星的辐射、黄道光、远距离恒星的辐射、星云的辐射等均是偏振的. 产生天体辐射偏振的原因很多，有固体表面反射、微粒散射、分子散射、自由电子散射、磁场中束缚电子的共振辐射、回旋共振辐射、同步加速辐射等. 测量天体辐射的偏振状态，可以为了解天体的物理状况、研究天体辐射机制和天体辐射经过的介质特性提供重要信息.

理论天体物理学(theoretical astrophysics) 天体物理学的一个分支学科. 它根据理论物理学的原理和方法以及实测天体物理学提供的天体观测资料来研究天体的物理性质和物理过程. 它和理论物理学结合密切，相互促进. 一方面用理论物理学方法解释天体物理现象；另一方面根据宇宙天体所具有的特殊物理条件来探索物理规律，发展理论物理学. 辐射理论、原子核理论、引力理论、等离子体理论、基本粒子理论、凝聚态理论是理论天体物理学的理论基础. 恒星大气理论、元素合成理论、恒星内部结构理论等是理论天体物理学中比较成熟的部分. 按照研究内容和所应用的理论，它又可分为高能天体物理学、相对论天体物理学、等离子体天体物理学、宇宙磁流体力学、宇宙气体动力学、中微子天体物理学等. 理论天体物理学既是理论物理学用于天体问题研究的应用学科，又是利用天体现象探索基本物理规律的基础学科.

恒星大气理论(theory of stellar atmosphere) 理论天体物理学的一个分支学科. 它根据对观测得到的恒星光谱的解释，研究恒星大气的物理状态(例如温度、密度、压力)随大气内不同深度的分布规律，研究恒星大气内原子的激发和电离、辐射的吸收和发射等物理过程，以及恒星大气的化学组成等. 恒星大气理论是在恒星大气处于辐射平衡和局部热动平衡的前提下建立起来的，现在已建立起各种光谱型的恒星大气模型. 此外，现代的恒星大气理论还研究恒星大气的对流、湍动、自转、质量外流、电磁过程和非热现象，以及偏离辐射平衡和局部热动平衡的恒星大气、化学组成反常的恒星大气等. 研究的主要内容有：建立恒星大气模型；利用恒星大气模型计算出的恒星连续光谱能量分布理论曲线，与实测的恒星光谱比较分析；研究恒星光谱的吸收线，确定恒星大气的化学组成、物理状态和物理过程；研究恒星光谱的发射线等.

恒星大气模型(model of stellar atmosphere) 恒星大气理论的一个基本课题. 内容是建立描述恒星大气内各种物理量(例如温度、密度、气体压力、电子压力等)随大气深度的分布规律的理论模型. 在假定恒星大气处于辐射平衡和局部热动平衡及流体静力学平衡的前提下，根据给定的恒星表面的有效温度、表面重力加速度和大气化学组成，用逐次近似方法求解温度、气体压力、电子压力与大气深度之间的关系式，逐步得出恒星大气物理量随大气深度的变化规律. 目前，对一些著名的恒星，例如天琴座 α 星(中文名“织女一”)、天鹅座 α 星(中文名“天津四”)、大犬座 α 星(中文名“天狼”)、小犬座 α 星(中文名“南河三”)以及太阳，已建立各自特有的大气模型.

流体静力学平衡(hydrostatic equilibrium) 天体物理学的一个概念. 指恒星大气中的气体和辐射的向外压力(气体压和辐射压)与向内的引力之间的平衡. 此种平衡关系用流体静力学方程给出为

$$\frac{dP_g}{dh}+\frac{dP_r}{dh}=\frac{GM\rho}{h^2},$$

式中 P_g 和 P_r 分别为气体压力和辐射压力，h 为线深度，M 为物质质量，ρ 为物质密度，G 为万有引力常数. 由此平衡方程，再连同压力的定义式以及由已知化学组成和热动平衡关系得出的电子密度的公式，三个关系式联立，运用数值积分法，就能得出恒星大气的各个物理量(温度、密度、气体压力、电子压力等)随大气深度的分布的理论计算值.

辐射平衡(radiative equilibrium) 研究恒星大气的基本理论假设. 此假设认为：在稳定的恒星大气中，每一体积元内在单位时间内放出的总辐射能 E_- 和单位时间内吸收的辐射能 E_+ 相等，即 $E_-=E_+$，或

$$\int_0^\infty\int_0^{4\pi} j_\nu d\omega d\nu=\int_0^\infty\int_0^{4\pi} I_\nu\chi_\nu d\omega d\nu,$$

式中 j_ν 为辐射系数，为单位质量的物质、在单位时间内、在频率区间 $d\nu$ 内、在单位立体角 $d\omega$ 内所辐射的能量；I_ν 为辐射强度，为在单位时间内、垂直经过面积元 $d\sigma$、通过单位立体角 $d\omega$、在频率区间 $d\nu$ 的能量；χ_ν 称为吸收系数，定义为

$$dI_\nu=-I_\nu\chi_\nu\rho ds,$$

式中 dI_ν 为强度 I_ν 的辐射垂直投射于厚度为 ds 的吸收物质层后得到的增量(负值),ρ 为吸收物质的密度.

局部热动平衡(local thermodynamic equilibrium) 描述恒星大气状态而引入的一种假设. 由施瓦茨席尔德(Schwarzschild,K.)于 1921 年和米尔恩(Milne,E. A.)于 1906 年分别提出. 恒星大气实际上不处于热动平衡状态. 对于一些与热动平衡偏离不大的恒星大气,假定在其局部区域里的物理参量遵从热动平衡下的物理规律,在这个局部区域内可以引入温度来表征它的热状态,可以应用热动平衡下导出的普朗克定律、麦克斯韦定律、玻耳兹曼公式和萨哈公式等来研究热辐射规律、电子的速度分布、原子的激发和电离状态等. 把根据局部热动平衡假设建立起恒星大气理论所得出的结论和观测结果比较,表明这种假设基本上接近恒星大气的物理状态.

萨哈公式(Saha formula) 一个天体物理学公式. 是表述在热动平衡状态下,单位体积内在不同电离级的原子数的天体物理学公式. 1920 年由印度天文学家萨哈(Saha,M. N.)导出. 假设在 $1cm^3$ 内有 n_r 个和 n_{r+1} 个处在 r 次电离和 $r+1$ 次电离的原子数,则萨哈公式为

$$\frac{n_{r+1}}{n_r}=\frac{u_{r+1}}{u_r}\cdot\frac{2(2\pi m_e)^{2/3}(kT)^{5/2}}{P_e h^3}\cdot\exp\left(-\frac{\chi_r}{kT}\right),$$

式中 m_e 为电子质量,h,k 分别为普朗克常数和玻耳兹曼常数,T 为绝对温度,χ_r 为 r 次电离原子的电离电势,u_{r+1},u_r 分别为 $r+1$ 次电离原子和 r 次电离原子的配分函数.

辐射转移理论(theory of radiative transfer) 研究天体上能量传递的一种理论. 用以研究热辐射通过既有吸收又有发射的介质时的变化情况和热辐射与物质的相互作用的辐射转移规律. 恒星大气中能量的传递基本上是靠热辐射方式进行的. 主要是建立辐射转移微分方程并求解. 在介质对辐射流的吸收与辐射频率无关的假设下,在辐射平衡的条件下所得出的辐射能转移方程的解是第一次近似,用逐次近似的方法可以研究吸收与频率有关的不同光谱型的恒星大气的辐射转移.

灰色大气(grey atmosphere) 一种理论模型. 所谓灰色大气,指假设大气的连续吸收系数与波长无关的一种模型大气,是实际恒星大气的近似描述. 它虽是早期模型,现仍广泛采用. 由此能算出大气温度的分布规律. 但是灰色大气不能解释观测到的恒星连续光谱能量随波长的分布,更无法解释此种分布的跳变特征.

不透明度(opacity) 天体物理学术语. 是表征介质对穿过其中的辐射吸收能力强弱的量. 辐射通过介质要受到介质的吸收,介质对辐射吸收愈强,不透明度愈大. 造成恒星介质不透明的吸收过程对于不同波段、不同温度、不同层次是不同的. 例如,在可见光波段,造成恒星大气连续吸收的主要是原子的光致电离和电子的自由-自由跃迁和自由电子的散射等.

线吸收(line absorption) 天体物理学的一个概念. 原子从低能级 E_1 跃迁到高能级 E_2 时,就要吸收频率为

$$\nu=\frac{E_2-E_1}{h}$$

的光量子(h 为普朗克常数),该频率处的辐射减弱,形成吸收线,这种吸收称为线吸收.

非相干散射(incoherent scattering) 一种散射过程. 指再辐射的光量子频率与所吸收的光量子频率不相等的散射过程称为非相干散射(参见“相干散射”).

相干散射(coherent scattering) 一种散射过程. 指再辐射的光量子频率与所吸收的光量子的频率相等的散射过程. 原子因吸收外界光量子的能量而被激发到高能态,但它又会回到原来的低能态而放出光量子. 相干散射中这些光量子虽然频率相等,但再辐射的光量子向各个方向散射,仍在该频率上形成吸收. 如果再辐射的光量子与所吸收光量子频率不相等,则散射称为非相干散射. 在天体物理现象中经常遇到的是非相干散射.

汤姆森散射(Thomson scattering) 一种散射过程. 指由于自由电子(或束缚电子)引起电磁辐射的散射. 根据经典电动力学,当任一频率的外来电磁辐射投射到自由电子(或束缚电子)上时,电子在电磁辐射作用下,以同样的频率作强迫振动(或在原来振动上叠加强迫振动),同时不断向各个方向发出该频率的次波. 这种现象由英国物理学家汤姆森(Thomson,J.)于 1903 年发现,故名. 当频率低于 10^{20}Hz 时,自由电子的汤姆森散射截面(即一个自由电子的散射系数)等于 $6.65\times10^{-25}cm^2$.

瑞利散射(Rayleigh scattering) 一种散射过程. 指当任一频率的光投射到线度小于光波波长的微粒上,微粒将向各个不同方向发出相同频率的次波,这种散射是前后光波频率不改变的光散射,因 1927 年英国物理学家瑞利(Rayleigh,L.)首先研究而得名. 瑞利散射的规律是:

1. 散射光的强度与入射波的频率的 4 次方成正比,短波长的光受到更强烈的散射;

2. 散射光强度与入射方向有关,散射光强度正比于$(1+\cos^2\theta)$,其中 θ 是散射光和入射光方向之间的夹角.

恒星大气中氢原子(具有束缚电子)的瑞利散射

是连续吸收的源泉,能有效地散射太阳型恒星的紫外辐射.氢原子的散射截面(一个氢原子的散射系数)为

$$\frac{8\pi}{3}\left(\frac{e^2}{m_e c^2}\right)^2\left(\frac{1026}{\lambda}\right)^4=6.65\times10^{-25}\text{cm}^2,$$

式中 λ 为波长,c 为光速,e 和 m_e 为电子电荷和质量.

拉曼散射(Laman scattering) 一种散射过程.是透明气体、液体、固体介质的分子对入射光的一种特殊的散射现象.因印度物理学家拉曼(Laman,C.V.)首先发现而得名.如果入射光是单色光,则在散射光谱中,在原有谱线两侧的对称位置上出现一些新的弱谱线,长波侧的谱线较短波侧的谱线强些.二者统称为拉曼谱线.产生拉曼散射的原因是散射分子的转动能态和振动能态发生变化,结果使得散射光子频率不同于入射光子的频率.

复合线(recombination line) 天体物理学术语.指自由电子被质子或原子核俘获复合而成为原子,在复合过程中电子放出能量子(光子)在光谱中形成的发射线.如在气体星云中,由于物质密度和辐射密度极低,绝大多数中性氢原子处于基能态,而处于激发态的原子很少.氢原子常吸收紫外光子而电离形成自由电子和质子.自由电子在运动一段时间后,又会被质子俘获而形成中性原子,这时电子或者直接复合到基态放出紫外光子,在光谱紫外区形成发射线(莱曼线系),或者先复合到某一激发态,再向下逐级跃迁,最后回到基态,同时放出光子,在光谱中形成巴耳末线系的发射线.

禁线(forbidden line) 天体物理学术语.指光谱中由于禁戒跃迁产生的谱线.原子在各能态间跃迁应遵从选择定则,不满足选择定则的跃迁概率很小,故称禁戒跃迁.只有原子的高能态是亚稳态时禁戒跃迁容易发生.在气体星云和日冕的光谱中可以观测到禁线.

辐射阻尼(radiation damping) 谱线致宽机制中主要的一种.指因辐射引起发射体系的运动衰减而导致谱线变宽.按照经典电动力学理论,发射(吸收)光的原子可看做为谐振子.当谐振子自由振动时,辐射出的电磁波带走能量,振动会衰减下来,所以把这种现象看做振动被一种消耗能量的阻尼力所致,这个阻尼力即称为辐射阻尼.由于振动受到辐射阻尼的作用,所辐射的电磁波振幅不断衰减,这样就得到具有一定宽度的谱线.按照量子力学理论,谱线是由原子分立的能级间的跃迁引起,原子在各能级上只能停留一有限的时间 Δt,根据能量和时间间的测不准关系($\Delta E\approx h/\Delta t$),能级不是无限窄的,而是有一定宽度的,因此,两个有一定宽度的能级跃迁形成的谱线便有一定宽度,称为辐射致宽.

辐射致宽(radiation broadening) 见“辐射阻尼”.

碰撞阻尼(collisional damping) 亦称碰撞致宽.谱线致宽机制中的一种.指辐射振子受到粒子碰撞引起谱线致宽.每次碰撞,振子的振动停止一瞬间,马上又激发新的具有任意的相位和振幅的振动,因此,碰撞阻尼使谐振子发出一段一段有限长的波,长度由两次碰撞的间隔时间决定.碰撞减短了所辐射的波的顺序性,增加了谱线的宽度.

碰撞致宽(collisional broadening) 即“碰撞阻尼”.

多普勒致宽(Doppler broadening) 天体物理学术语.指恒星大气中的原子、分子、电子作不规则的运动导致的谱线变宽.由于多普勒效应,以各种不同速度运动的粒子所辐射的谱线产生不同大小的位移,总的效果使谱线变宽.谱线变宽的程度和粒子的运动速度有关,高温时谱线增宽很大,平均说来,氢线比重元素的谱线多普勒致宽程度大.

湍动致宽(turbulence broadening) 天体物理学术语.指由湍动运动引起的谱线致宽.这种机制由罗斯兰德(Rosseland,S.)于 1928 年提出.恒星大气中的原子除作热运动外,还参与大气内的湍动,或者是大的气团的流动,或者是原子以高于热运动速度的速度运动.湍动是比原子的尺度大得多而比恒星半径小得多的气体团的运动.通常把比光子自由程大的湍动称为宏观湍动,比光子自由程小的湍动称为微观湍动.因湍动产生的多普勒效应会使谱线产生红向和紫向的位移造成谱线变宽,故名.湍动致宽可以部分地看做气团的热动多普勒致宽.在许多恒星光谱中,都存在因湍动而出现的谱线致宽,特别对超巨星更加明显.

斯塔尔克效应(Stark effect) 一种物理现象.指在强电场作用下原子发射的谱线分裂成几条的现象.由德国物理学家斯塔尔克(Stark,J.)于 1913 年发现,故名.对于氢原子在电场强度不很大时,原子能级的分裂宽度和外电场强度成正比.在外电场强度很强时(超过 10^5V/cm),分裂宽度和外电场强度的平方成正比.恒星大气中含有大量的离子和电子,它们相互接近时产生很强的微观电场而引起斯塔尔克效应,总的效果是使谱线变宽.

塞曼效应(Zeeman effect) 一种物理现象.是放置于强磁场(约零点几至几个特斯拉)中的光源发射的谱线,受磁场影响而分裂成几条的现象.1896 年,由荷兰物理学家塞曼(Zeeman,P.)发现,故名.当磁场方向与视线方向平行时,谱线分裂为两条;当磁场方向与视线方向垂直时,谱线分裂为三条.这是由于外磁场和原子磁矩发生作用使原子能级发生分裂.能级分裂的大小和外磁场强度成正比.根据谱线

分裂宽度可以测定天体磁场强度.

压力致宽(pressure broadening) 天体物理学术语.指由于恒星大气内原子相互碰撞而改变原子辐射频率导致谱线变宽.在电场中,原子能级分裂成几个次能级,跃迁时产生几条细小的子谱线,也使谱线变宽.这两种效应都与压力有关,有时统称为压力致宽.

振子强度(oscillator strength) 天体物理学术语.按照经典电动力学理论,电磁波的发射和吸收是由谐振子作简谐振荡引起的.在给定的谱线内,一个原子的吸收作用可以用振子数来表示,称振子数为振子强度.它表示处在某一能态和某一体积中的某一元素参与吸收的相对原子数目.

热辐射(thermal radiation) 一种电磁辐射.指物态处于热动平衡状态(物质粒子的能量分布遵守波耳兹曼分布律)下因自身温度而发射出的电磁辐射.热辐射的能量随频率的分布遵守普朗克定律:

$$B_\nu(T)=\frac{2h\nu^3}{c^2}\cdot\frac{1}{e^{h\nu/kT}-1},$$

式中 $B_\nu(T)$ 为绝对黑体的辐射强度,T 为绝对温度,ν 为辐射频率,c 为光速,h 为普朗克常数,k 为波耳兹曼常数.

非热辐射(non-thermal radiation) 一种电磁辐射.指物态偏离热动平衡状态很大的辐射源所发出的辐射.有些天体,如类星体、中子星、X 射线源、γ 射线源等的辐射的谱形、偏振等与热辐射有明显区别,例如,类星体和射电星系发射有偏振特性的幂律型射电频谱.非热辐射的形成机制有多种,如相对论性电子在磁场中的同步加速辐射(类星体、射电星系)、相对论性电子在等离子体中运动时的切仑柯夫辐射(太阳Ⅱ型及Ⅲ型射电爆发)、天体微波激射源放大(星际分子射电源)等.

同步加速辐射(synchrotron radiation) 一种非热辐射.指速度接近光速的高能电子在外磁场中沿圆轨道或螺旋轨道运动时产生的辐射.因最早在电子同步加速器中发现而得名.现在认为许多天体的非热辐射即为同步加速辐射,例如,射电星系的射电辐射、太阳射电爆发等.这种辐射机制具有辐射功率强、有显著偏振、方向性强等特点.

切仑柯夫辐射(Cerenkov radiation) 一种非热辐射.指高速带电粒子在介质中穿行时,当粒子速度大于光在介质中的相速度而产生的一种辐射.它是一种包括可见光在内的连续谱,具有明显的方向性和强偏振等特点.由苏联物理学家切仑柯夫(Черенков, Л. А.)和瓦维洛夫(Вавилов, С. И.)于1934年首先发现.作匀速直线运动的带电粒子,当其速度大于介质中的光速时,它辐射的电磁波将集中在粒子后方的一个圆锥形区域之中,粒子位于圆锥的顶点上.类似于船的航行速度超过水平波的传播速度时,在船首附近形成的V形波.切仑柯夫辐射实质上是一种电磁"冲击波"现象.太阳Ⅱ型和Ⅲ型射电爆发,可以用相对论性电子在等离子体中穿行时的切仑柯夫辐射来说明.

曲率辐射(curvature radiation) 一种非热辐射.指高能电子沿强磁场磁力线方向运动时,由于磁力线本身的弯曲而产生的电磁辐射.在强磁场中运动的高能电子,当磁场不平行于电子运动速度时,会引起同步加速辐射;但当磁场很强,任何横越磁力线的运动实际上不可能存在,电子只能沿磁力线运动.在磁力线弯曲的情况下,高能电子沿磁力线作曲线运动,因而也会产生辐射,即曲率辐射.曲率辐射的谱形特点和同步加速辐射相似,例如,频谱功率在低频段正比于频率的立方根,在高频段按指数衰减,曲率辐射常用来说明脉冲星、类星体等高能天体的辐射现象.

回旋加速辐射(cyclotron radiation) 一种非热辐射.指电子在外磁场中沿圆轨道或螺旋轨道运动时所产生的辐射.这种辐射的辐射功率小于同步加速辐射,具有单色性强、偏振等特性.某些天体辐射(例如太阳耀斑、中子星)的X射线的线状发射可用这种辐射来说明.

韧致辐射(bremsstrahlung) 一种非热辐射.指高能带电粒子在与原子核相碰撞突然减速时产生的辐射.在天体物理中,有些X射线源的辐射可用韧致辐射来说明.

康普顿散射(Compton scattering) 一种物理效应.指短波电磁波(例如X射线、γ射线)射入物质被散射后波长变长的现象.于1922—1923年由美国物理学家康普顿(Compton, A. H.)和中国物理学家吴有训发现.其物理机制是高能光子被物质中心电子散射时,将一部分能量传递给电子,从而使光子波长变长.

逆康普顿散射(inverse Compton scattering) 一种物理过程.是康普顿散射的逆过程,即高能电子与低能光子碰撞而使低能光子获得能量的散射过程.宇宙空间和天体中普遍存在各种低能光子,由于逆康普顿散射可使这些低能光子变成高能光子,这种过程可用于解释宇宙X射线的起源.

恒星内部结构理论(theory of stellar interior structure) 理论天体物理学的一个分支学科.它根据观测所得的恒星质量、光度、半径、表面温度和化学成分等数据,利用物理理论来研究恒星内部的物理情况.在恒星是一稳定的气体球、处于流体静力学平衡的假定下,研究恒星内部的温度、密度、压力等物理参量随半径方向的分布情况;确定恒星的化学组成和元素的分布情况;确定恒星发光的能源及传

递能量的过程等. 例如，对于太阳，利用恒星内部结构理论，得到太阳中心温度为 1.5×10^7K，中心密度为 160 000kg/m^3，内部能源质子-质子反应产生的能量占总能量的 90%以上，小部分能量来自碳氮循环，化学成分按质量计：氢为 71%，氦为 27%，其他重元素为 2%等.

多方球(polytrope)　一种恒星模型. 指在自引力作用下不均匀的和可压缩的内部物质达到平衡的一种恒星模型. 其中压力与密度的$(1+1/n)$次方成正比，n 称为多方指标，$(1+1/n)$称为多方指数.

质子-质子反应(proton-proton reaction)　一种物理过程. 指由四个氢核聚变成一个氦核的一系列热核反应过程. 为小质量、低光度的主序星的主要产能机制，它由三个分支反应过程分别完成. 太阳内部每秒发生 10^{38} 次此种反应，太阳所辐射的能量的 90%以上是质子-质子反应过程提供的.

碳氮循环(carbon nitrogen cycle)　一种物理过程. 指由碳、氮起触媒作用，促使四个氢核聚变为一个氦核的热核反应过程. 大质量、高光度的主序星的能量主要由碳氮循环产生，它由两个分支反应过程分别完成.

元素合成理论(theory of nucleosynthesis)　解释宇宙中各种元素及其同位素的形成过程的理论. 英国伯比奇夫妇(Burbidge, G. and Burbidge, E. M.)、霍伊尔(Hoyle, F.)和美国福勒(Fowler, W. A.)提出所有元素都是在恒星中合成的学说，认为各种元素及其同位素都是由氢通过恒星中发生的八个过程(氢燃烧、氦燃烧、α 粒子与 ^{20}Ne 反应过程、原子核热动平衡过程、慢中子俘获过程、快中子俘获过程、质子俘获过程以及生成氘、锂、铍、硼轻元素的 x 过程)逐步合成后由恒星抛射到宇宙空间的. 后来又提出一些新的过程(碳燃烧、氧燃烧、硅燃烧)及氦由宇宙大爆炸时生成的等学说.

元素丰度(abundance of element)　天体物理学术语. 指宇宙中各种天体的各种元素及其同位素的数密度的相对值. 通常取硅的丰度为 10^6，其他元素的丰度按比例确定. 常用表或图来表示. 以元素质量数为横坐标，以丰度为纵坐标，所得到的曲线称为元素的丰度曲线. 它反映出元素按质量数的分布规律.

p **过程**(p-process)　一种假设过程. 指俘获质子形成富质子重原子核合成的一种假设过程：低于铈的原子核(其中中子分离能量较高)，可从中子俘获过程的产物出发，再通过俘获质子放出光子的反应而生成；高于铈的原子核(其中中子分离能量较低)，则从中子俘获过程产物出发，可再通过吸收光子放出中子的反应而生成. p 过程可能发生在温度高于 10^9K 和密度小于 10^7kg/m^3 的超新星的壳层中. p 是英文 proton(质子)的第一个字母.

r **过程**(r-process)　一种俘获过程. 指快的中子俘获过程. r 是英文 rapid(快速)的第一个字母. r 过程的时间约为 10～100s，在此过程中，某些原子核的中子俘获寿命短于 β 衰变寿命，生成的原子核尚未 β 衰变就又俘获新的中子，在俘获过程过去以后，富中子元素才能通过 β 衰变变成 β 稳定元素和富中子同位素. 比铋重(直到核子数目≈298)的所有元素和比铁重的富中子同位素，都可由 r 过程产生. 在超新星爆发时可能发生此过程.

s **过程**(s-process)　一种俘获过程. 指慢的中子俘获过程. 两次俘获之间的时间约为 10～100a. s 是英文 slower(较慢)的第一个字母，在此过程中，铁峰元素由于在弱中子流中逐步俘获自由中子而合成稳定的富中子的重元素核，因而在俘获另一个中子以前来得及发生 β 衰变，此过程是缓慢而可靠的元素合成过程. 发生在恒星演化的红巨星阶段壳层之间的区域，此区域的密度高达 10^8kg/m^3，温度约 3×10^9K，而中子密度假定为 10^{16}个/米^3. s 过程可能形成直至核子数目为 208 的稳定核素，因此进一步俘获中子会立即导致 α 衰变而恢复成铅或铊，而至此为止. s 过程的中子来源最可能是与氦核耗尽后在双壳层燃烧期间氦的热不稳定性有关. 当恒星质量小于 9 倍太阳质量时，可能发生此过程.

e **过程**(e-process)　一种统计平衡过程. 指一种假设的合成铁峰元素的一系列核反应统计平衡过程. e 是英文 equilibrium(平衡)的第一个字母. 当温度大于 5×10^9K 和密度大于 3×10^9kg/m^3 时，高能光子与原子核发生大量碰撞，碰撞导致核的碎裂，而碎片又很快和其他粒子结合，因而在碎裂与形成之间建立统计平衡. 因铁峰元素具有最大的结合能，所以粒子最终被这些原子核俘获. e 过程可能出现于超新星爆发过程中.

宇宙气体动力学(cosmic gasdynamics)　天体物理学的一个分支学科. 即把气体动力学方法应用于天文学研究的学科. 其任务是应用气体动力学方法，深入解释天体物理现象和探讨天体运动规律. 已广泛用于研究太阳，预言太阳风的存在，研究太阳风在行星际空间中的作用、太阳的结构和动力学问题；研究星际物质的运动特征；研究恒星的形成和演化；研究星系，发展旋涡结构的密度波理论，解释气体星云坍缩成星系的过程；研究类星体的双源结构；宇宙早期演化过程中气云块的碎裂过程、星系和星系团的形成等.

金斯不稳定性(Jeans instability)　由万有引力产生的一种不稳定性. 20 世纪初，英国金斯(Jeans, J. H.)最早提出，故名. 一个自引力体系，若基态是均匀或大致均匀的，则存在一个临界波长(常称金斯

波长,一般约为 10^{20}cm),尺度大于此临界波长的密度扰动将随时间增长,即密度大的部分将变得更密,成为引力不稳定性. 一般只有转动体系才可能出现金斯不稳定性. 但对基态是非均匀的体系,也应用金斯不稳定性作定性研究.

瑞利-泰勒不稳定性(Rayleigh-Tayler instability) 一种流体动力学不稳定性. 是天体物理中常见的现象. 在加速度场中,密度较大的流体位于势能较高的位置,而密度较小的流体位于势能较低的位置. 两种流体分界面上一定的小扰动会形成不稳定性,使不同密度的流体层混合才达到整个系统的稳定.

宇宙磁流体力学(cosmic magnetohydrodynamics) 理论天体物理学的一个分支学科. 是研究导电流体与电磁场相互作用的磁流体力学在天体物理中的应用. 宇宙中普遍有磁场,宇宙中大部分物质都处于气体和等离子体状态,等离子体在磁场中的运动是天体物理学的研究内容之一,磁流体力学则是一个重要的研究工具. 宇宙磁流体力学的特点是:以流体力学和电动力学为基础,把流场方程和电磁场方程联立起来,并且引进许多新的特征过程;研究对象的特征长度一般相当大,因此电感的作用远大于电阻的作用;有效时间非常久,虽然由电磁原因引起的某些作用力不大,但能产生重大效应. 主要研究:在介质流动下宇宙磁场的演变;磁流体力学波的情况;磁场对平衡的不稳定性,层流转换为湍流的不稳定性,热力学不稳定性的影响和磁场与湍流的相互作用.

阿尔文波(Alfven wave) 一种磁流力学波. 沿磁力线运动的磁流体切变波,是振动方向与传播方向垂直的横波. 因系瑞典学者阿尔文(Alfvén,H.)所发现,故名. 在磁场中的高电导流体内,磁力线与物质耦合在一起,即"冻结"在等离子体中,当垂直于磁场方向受到一种局部扰动时,电离物质总是力图在它垂直于磁力线的运动中与磁力线一起运动,而磁力线以弹性恢复力的作用而反抗任何扰动,结果引起横向振荡,产生阿尔文波. 宇宙天体大多属带磁高电导等离子体,所以阿尔文波在天体等离子体现象中起重要作用.

磁声波(magnetosonic wave) 一种特殊的磁流力学波. 是磁流体力学中的声波与阿尔文波互相耦合形成的一种波. 一般指冷等离子体垂直于外磁场方向传播而扰动电场与外磁场及波矢都垂直的低频波. 从电场扰动看是横波,而粒子振荡近似地在波的传播方向,与声波相似,故名. 在热等离子体中,磁声波是指色散关系涉及磁压强及热压强的磁流体力学波. 传播方向可与磁场成任意角,以相速度的高低,有快磁声波与慢磁声波之分. 当温度趋于零时,只存在快磁声波.

磁流体力学湍流(magnetohydrodynamic turbulence) 一种运动过程. 指带磁导电流体质点互相混杂、迹线极不规则的运动. 湍流的速度和压力等的时间平均值有不规则的涨落,具有高雷诺数. 流体的动能和磁能在此种湍流的最后阶段,以粘滞和焦耳损耗方式转变为分子热能. 湍流使平均运动增加了很强的动态摩擦. 观测证明,此种湍流在天体物理现象中普遍存在,例如,太阳对流层、致密星的吸积盘、星系中的气盘、超新星遗迹所代表的激波波阵面后的区域、太阳风等,都有此种湍流出现.

电流片(current sheet) 亦称中性片. 天体物理学术语. 指在两个反向磁场之间的电流薄层. 等离子体从两侧流入薄层,因磁合并以及不稳定过程而加热和加速. 不稳定性使电流片溃散,其中磁场为零处产生强感应电场,使带电粒子加速.

等离子体天体物理学(plasma astrophysics) 理论天体物理学的一个分支学科. 它应用等离子体物理学理论和成果研究天体物理现象. 许多天体物态处于等离子态,远离热动平衡状态. 等离子体天体物理学着重研究天体等离子体中湍动状态的物理过程. 可用于研究射电源的结构、超新星遗迹、电离氢区、脉冲星、行星磁层、行星际物质、星际物质和星系际物质等.

等离子体湍动加速(plasma turbulence acceleration) 一种物理现象. 指由于微小扰动在等离子体中激起各种湍动元(等离子体波)与带电粒子碰撞引起能量交换,导致粒子加速的现象. 此种加速带有统计性质,碰撞总是导致粒子平均能量的增加. 湍动元尺度越小,加速效率就越高. 因为等离子体中存在各种小尺度湍动元,所以湍动加速效率非常高,比费米加速效率高得多. 如果等离子体波的相速度大于被加速粒子的热运动速度,则只能加速少数快粒子,称为等离子体湍动纯加速;如果波的相速度小于粒子的热运动速度,就能加速大多数粒子,称为等离子体湍动加速. 等离子体湍动加速已广泛用于解释太阳缓变射电、脉冲星、类星体和星系核等的粒子加速.

等离子体反常输运(anomalous transport in plasma) 一种物理效应. 指等离子体中由带电粒子与集体扰动场之间的某种等效碰撞引起的效应. 等离子体中带电粒子与带电粒子之间的作用是长程的库仑力. 由于德拜屏蔽效应,计算输运系数时,常可用短程的屏蔽库仑势,按两体碰撞理论来处理,此即经典输运理论. 反常输运则指不属于经典输运的输运过程,但引起的耗散比经典输运大得多. 当等离子体出现不稳定和湍流时,反常输运变得特别重要. 反常加热、反常扩散和反常热导等都与反常输运密切相关.

高能天体物理学(high energy astrophysics) 理论天体物理学的一个分支学科.它研究发生在天体上的高能现象和高能过程.20世纪60年代以来发现的类星体、脉冲星、X射线源、γ射线源、超新星爆发、星系核活动等高能天体和高能过程,要用高能物理学的规律来解释,因此,发展起高能天体物理学.主要研究成果有:发现若干中微子过程对恒星晚期的演化过程的重要作用;对中子星、超密态星体的物态和结构的研究,对宇宙X射线爆发、γ射线爆发和星系核爆发等的理论解释都取得重要进展.

中微子天体物理学(neutrino astrophysics) 理论天体物理学的一个分支学科.研究恒星上可能发生的中微子过程以及这些过程对恒星结构和演化的作用;研究原子核β衰变时,理论上预言存在一种不带电、静止质量为零的基本粒子即中微子.从1955年起用四氯化碳接收太阳核反应所发射出的中微子,实验结果比理论预期的要差一个数量级,即太阳中微子短缺,至今尚无圆满解释.理论研究表明:中微子对恒星演化晚期起重要作用,中微子带走大量能量会加速恒星演化过程,在超新星爆发和中子星形成过程中,中微子也起着重要作用.

电子对湮没中微子过程(electron pair annihilation neutrino process) 一种物理过程.指电子和正电子相互碰撞发生湮没而产生一对正反中微子的过程.丘宏义和莫里森(Morrison,P.)于1960年首先进行研究.这是一个通过中介玻色子传递的弱作用过程.每次湮没,恒星将损失1MeV以上的能量.当恒星温度在10^9K以上时,此过程是恒星能量损耗的主要过程,对恒星的演化有重要作用.

光生中微子过程(photoneutrino process) 一种物理过程.指光子与电子碰撞产生一对正反中微子的过程.丘宏义和斯特布勒(Stabler,R.C.)于1961年首先进行研究.这是一种电子吸收光子的电磁作用和由中介玻色子传递的弱作用两者组合起来的过程.由于中微子有很强的穿透力,能迅速逃逸,所以上述过程使体系的能量被中微子迅速带走.能量损耗约正比于温度的8次方,在恒星密度小于10^8kg/m^3、温度低于4×10^8K的范围内,光生中微子过程是主要过程.

等离子体激元衰变中微子过程(plasmon decay neutrino process) 一种物理过程.指等离子体中各种形式的波的量子(即等离子体激元,可看做准粒子)衰变为一对正反中微子的过程.亚当斯(Adams,J.B.)等人于1963年首先进行研究.是由等离子体激元湮没为正负电子对的电磁作用和由中介玻色子传递的弱作用两者组合起来的过程.由于中微子有很强的穿透力,能迅速逃逸,所以上述过程使体系的能量被中微子迅速带走,成为恒星能量损耗的主要过程,对恒星的演化有重要作用.

乌卡过程(Urca process) 一种物理过程.指产生中微子的一种过程.主要发生在铁族元素的一系列核反应中,反应过程伴随着中微子的很高的形成率,被认为是恒星坍缩的一种原因.由于中微子能很快地不被看见地携走能量,犹如巴西里约热内卢的乌卡赌场能很快地不被看见地携走钱财一样,故名.

宇宙线(cosmic rays) 一种微粒流.指来自宇宙空间由以光速运动的各种高能微观粒子构成的微粒流.在宇宙空间中穿行的称为初级宇宙线;通过地球大气、受大气作用的称为次级宇宙线.初级宇宙线由质子——氢原子核(约占87%)、α粒子——氦原子核(约占12%)和少量的锂、铍、硼、碳、氮、氧的原子核,以及极少量的重元素原子核组成.初级宇宙线进入地球大气上层时,与氧原子和氮原子相碰撞,引起原子核的分裂,并产生其他粒子,这些粒子与周围物质作用以及自身的转变,形成次级宇宙线.次级宇宙线可以到达地面并深入地下.宇宙中,宇宙线的发生源很多,例如超新星和银河系内的射电星云等,太阳上的质子耀斑是太阳发射的宇宙线源.宇宙线的增强能引起无线电通信的中断,并引起磁暴(称为宇宙磁暴).宇宙线与天体物理的研究和地球物理现象及日常生活都有密切关系.

费米加速机制(Fermi acceleration mechanism) 天体物理学的一个概念.指宇宙线粒子加速的一种可能的统计加速机制.由意大利物理学家费米(Fermi,E.)于1949年提出,故名.星际气体带动微弱磁场运动,带电粒子同磁场发生"碰撞",当磁力线与粒子相向运动,粒子就获得能量;当磁力线与粒子追赶碰撞,粒子就损失能量;粒子在多次碰撞后就成为高能的宇宙线粒子.此种机制用于天体物理、太阳物理和地球磁层研究.

相对论天体物理学(relativistic astrophysics) 理论天体物理学的一个分支学科.它应用广义相对论理论研究天体.主要研究内容包括:相对论宇宙学,研究宇宙的大尺度时空结构和几何特征;致密天体物理学,研究致密星,例如中子星、黑洞等天体;引力波天文学,研究各种天体过程的引力辐射对天体现象的影响;后牛顿天体力学,研究广义相对论对普通天体力学的各种修正.

引力收缩(gravitational contraction) 一种物理过程.指气体弥漫物质中密度高的区域因自身引力作用而收缩,使密度进一步增大的现象.一般认为太阳系、恒星、星系都是由于原始星云在自身引力作用下逐渐凝缩而成的.

克尔度规(Kerr metric) 天体物理学术语.指爱因斯坦引力场方程的一个轴对称解.由克尔

(Kerr,J.)于1963年得出,故名.它对应于电荷为零的轴对称旋转体在其外部所产生的稳定引力场.

施瓦茨席尔德度规(Schwarzschild metric) 天体物理学术语.指爱因斯坦引力场方程的第一个严格解.由德国天文学家和物理学家施瓦茨席尔德(Schwarzschild,M.)于1916年得出,故名.它表示无自转、球对称天体所产生的静态引力场的四维时空的度量性质.当离开球体相当远时,此度规即化为通常的闵科夫斯基度规.

罗伯逊-沃克度规(Robertson-Walker metric) 天体物理学术语.指适用于均匀和各向同性时空的一种度规.1935年和1936年分别由罗伯逊(Robertson,H.P.)和沃克(Walker,A.G.)证明,故名.适合上述要求的四维时空只有三种:体积有限的球状空间、体积无限的平直空间和体积无限的双曲空间.

引力波(gravitational wave) 亦称引力辐射.一种物理现象.即以波动形式传播的引力场.广义相对论预言,任何有加速度的质量将发出引力波.引力波具有在真空中以光速传播、具有能量、辐射强度极弱、穿透性极强等特点.目前,尚无直接实验证实,只有观测某些双星轨道参数变化的间接验证.

引力辐射(gravitational radiation) 即引力波.一种物理现象.美国科学家韦伯(Weber,J.)试图测量自由悬挂的一根大铝棒的振动来探测引力波,他曾宣布在银河系中心方向探测到了引力波,但迄今为止尚未为其他观测者所证实.近年来,通过探测天体运动状态的变化间接验证了引力波的存在.由于辐射阻尼的作用,双星系统将因损失能量而使公转周期变短.1978年12月,美国人泰勒(Tayler,H.D.)等人发现了脉冲星PSR1913+16的公转周期逐渐变短,周期频率为$(-3.2\pm0.6)\times10^{-12}$,这是引力波存在的间接定量证据.

光线引力偏转(gravitational deflection of light) 一种光线传播现象.指光线通过引力场时沿着弯曲空间的测地线传播的现象.根据广义相对论,一束光线经过大质量物体周围,光线将偏向该物体.例如,通过太阳边缘到达地球的星光,由于太阳引力场的作用将偏转1.75″.1919年,观测日全食时证实了广义相对论的这一预测.

引力透镜效应(gravitational lens effect) 一种天体物理现象.指引力场使光线偏转产生像透镜会聚光线那样的效应.广义相对论认为,光线沿引力场的弯曲空间的测地线传播.如果观测者到天体的视线上有一个大质量致密的前景天体,则遥远的天体的光经过大质量的前景天体会偏向,该大质量天体使光发生偏转,类似透镜对光线的会聚作用.于是,在真实位置两边可以生成两个像;如果起透镜作用的前景天体不是恰好在到遥远天体的视线上,可能生成几个像.1979年9月,沃尔什(Walsh,D.)等在美国基特峰天文台,用口径2.1m望远镜首次观测到,具有相同的发射线和红移($z=1.39$)的"类星体对"0957+561A和0957+561B,就是引力透镜效应造成的.后来又陆续发现1635+267,2345+007,1115+080,2016+112等有引力透镜效应的类星体,其中后两个类星体有三个像.

引力红移(gravitational redshift) 一种天体物理现象.指在引力场中的光源发出的光谱线向长波方向(红端)偏移的现象.偏移量与光源和观测者两处的引力势差的大小成正比.广义相对论认为:在引力场中,光振动要变慢,对应的波长变长.最初在白矮星(天狼星伴星)的光谱中发现这种现象.引力红移的大小与相对论的理论值相符合.

吸积(accretion) 一种物理过程.指天体以自身的引力把周围的气体、尘埃等物质吸引并积聚起来的过程.在太阳系形成和双星演化过程中吸积起重要作用.吸积过程中会释放能量.中子星或黑洞等致密星表面引力场很强,因此,被吸积的物质具有很大动能,会转化为波长极短的电磁辐射(如X射线、γ射线)发射出来.

脉冲星(pulsar) 一类天体.指发射短而强的无线电脉冲的特殊天体,是快速自转的中子星.1967年,由英国天文学家休伊什(Hewish,A.)和贝尔(Bell,J.L.)发现.它的特征是无线电脉冲(有的还有X射线、γ射线和可见光脉冲)周期非常稳定、短促而精确.其物质密度高达10^5kg/m^{3}以上;表面温度达10^8K,中心温度达6×10^{10}K;中心压力达10^{28}个标准大气压(1标准大气压=101 325Pa);磁场强度达10^8T以上;它的辐射能量平均为太阳辐射的100万倍,其大小为10km量级;距离约几千光年.现已发现600多颗.

中子星(neutron star) 一类天体.指主要由中子及少量的质子、电子组成的超密天体.密度达10^5kg/m^{3}以上,磁场强度达10^8T以上.它是处于演化晚期的恒星,是超新星爆发后的残核.超新星爆发时对其核心产生极大的压力,使物质原子的外层电子被挤进原子核内和质子结合,形成异常致密的中子态物质.中子星是已知体积最小的恒星,半径只有几十公里.蟹状星云中的脉冲星就是快速自转的中子星.

昌德拉塞卡极限(Chandrasekhar limit) 天体物理学的一个概念.指稳定白矮星的质量上限.恒星演化晚期,引力坍缩使中心密度增大,物质出现电子简并,简并压力与引力抗衡维持白矮星稳定的质量,称为昌德拉塞卡极限.20世纪30年代,由印裔美国天文学家昌德拉塞卡(Chandrasekhar,S.)精确计算

出该质量为 1.44 太阳质量.当质量小于1.44太阳质量,白矮星是稳定的;当质量大于 1.44 太阳质量,不能形成稳定的白矮星,星体会进一步坍缩.

黑洞(black hole)　理论预言的一种天体.是宇宙空间的强引力区域.因其脱离速度等于光速,不会有辐射发射出去,故名.黑洞的表面,即物质和辐射不能逸出的空间的边界称为视界.黑洞外的物质和辐射可以通过视界进入内部,而黑洞内的物质和辐射却不能穿出视界.在恒星演化晚期,星体质量超过 2 倍太阳质量时,在自引力作用下将坍缩成为一个黑洞.理论研究表明,在宇宙早期可能形成一些小质量(1^{10}t、半径为10^{-15}m)的小黑洞,其温度高达1.2×10^{12}K,有很强的辐射.此外,在有些星系的核心可能存在质量为 $10^8\sim10^{10}$倍太阳质量、半径为 2～200 天文单位的大型黑洞.

彭罗塞过程(Penrose process)　一种理论上的物理过程.指从电荷为零、轴对称的自转黑洞取出转动能的过程.由彭罗塞(Penrose,R.)于 1969 年首先提出,故名.若粒子以与黑洞转向相反的方向旋入黑洞的能层,并且在能层中分裂成与运动方向相反的两个碎粒,则与黑洞转向相同的碎粒能逸出黑洞,而且其能量大于原来旋入粒子的能量,超出的能量即取自黑洞的转动能.

奥本海默极限(Oppenheimer limit)　天体物理学的一个概念.指中子星的质量上限.一颗热核能源耗尽的星体,如果超过这个极限质量,不可能成为稳定的中子星.1939 年,美国物理学家奥本海默(Oppenheimer,J.R.)等证明稳定的中子星存在一个临界质量,故名.这个质量上限现在一般取 2 个太阳质量.质量大于奥本海默极限的恒星可能成为黑洞或其他致密星.

白洞(white hole)　理论预言的一种天体.与黑洞相反,白洞只能由中心向外发射物质和能量,而不能吸收外部区域的任何物质和辐射.白洞目前只是一种理论模型,尚未被观测所证实.

太阳物理学(solar physics)　天体物理学的一个分支学科.它应用天体物理学的理论和方法研究太阳的物理性质、化学组成、活动规律以及内部结构和演化;根据太阳的光度、质量、半径、表面温度等大量实验资料,建立太阳大气模型,即描述太阳大气的温度、密度、压力等随深度的分布规律;应用物理学理论探讨太阳内部结构、能量来源和传输过程以及太阳各层次(光球、色球、日冕)和活动中心(黑子、耀斑、冕洞等)的物理性质等.

20 世纪 70 年代发现太阳中微子的实测值与理论值相差很大,对目前太阳物理学理论提出了挑战.近年来发展起来的日震学研究,为太阳内部结构的研究提供了新的途径和方法.

太阳(sun)　一个天体.即太阳系的中心天体.太阳系的九大行星及其他天体均在其引力控制下绕其运转,是银河系中一颗恒星.它位于银道面以北约 8pc,离银心约 10000pc 的银河系旋臂中,离地球平均距离为 149597870km,半径为 695992km,质量为 1.989×10^{30}kg,角半径为31′59.3″;为一高热气体星球,表面温度为 5770K,中心温度约为 1.5×10^7K,总辐射功率(光度)为3.826×10^{26}W,目视星等为−26.74等,绝对目视星等为+4.83 等,表面重力加速度为 2.74×10^2m/s^2,表面逃逸速度为 617.7km/s;平均密度为 1400 kg/m^3,自转周期为 26.9d(赤道)、31.1d(极区),中心密度为 1.6×10^6kg/m^3,中心压力为 3.4×10^{16}N/m^2;化学组成中氢占 71%,氦占 27%,其他元素占 2%.

太阳大气可分三层:光球、色球和日冕,太阳大气中存在各种活动现象,如太阳黑子、日珥、耀斑、射电暴和 X 射线暴等,它们与地磁暴、极光、电离层突然骚扰等地球物理现象有密切关系.太阳中心区域进行氢核聚变反应,维持着太阳的巨量辐射.太阳的年龄为 50 亿年,据推算,其寿命约为 100 亿年.

光球(photosphere)　天体物理学术语.指肉眼观测到的太阳表面层,是太阳大气层最里面的一层,厚度约 500km.由它把太阳核反应区产生的能量直接发射到宇宙空间,地球上接收到的太阳能量基本上就是由光球发射出的.整个光球是明亮的,但各部分亮度很不均匀.在非扰光球中布满米粒组织;在光球活动区中,则有太阳黑子、光斑和白光耀斑等太阳活动现象出现.

光球实际上是沸腾的太阳对流层的顶部,它的温度随高度而不同,从里向外递减,平均温度为 5770K,在与色球交界处,降到最低值,约为 4300K(交界处以外又逆升,在日冕中高达 1.2×10^6K).光球物质密度约为 $8\times10^{-5}\sim4\times10^{-4}$kg/m^3,气体压力约为 10^{-4}N/m^2.光球里各种化学元素都处于中性原子状态或低电离度的离子状态.

太阳黑子(sunspot)　太阳活动最明显的特征.指太阳光球层上出现的巨大旋涡状气流.温度比光球低 1000～2000K,与光球相比,成为暗淡的黑斑,故名.黑子有强达 0.1～0.4T 的磁场.一个发展完全的黑子,由本影和半影两部分组成.它的形状不很规则,以圆形或椭圆形居多;有时单个出现,有时成对或成群出现,复杂的黑子群由几十个大小不等的黑子组成,它们在太阳表面不断地移动、发展,经过一段时间自行消逝.它的寿命长短不一,多数出现几天到几十天,短的只有几小时;少数长的,能够维持几个月,甚至一年半.小黑子的直径约为 1000km,而大黑子直径可达 2×10^5km,几乎所有的黑子都分布在日面纬度±40°的范围内,但赤道两旁±8°范围

内很少有黑子出现.

太阳黑子数呈周期性增减,周期平均约为 11a. 大黑子出现,地球上常发生许多地球物理现象,如磁暴和电离层暴,在两极地区引发极光等.

太阳黑子相对数(sunspot relative number) 亦称沃尔夫数,或沃尔夫相对数. 表示太阳黑子活动程度的一种指数. 由瑞士天文学家沃尔夫(Wolf,J. R.)于 1849 年提出. 其公式为

$$R=K(10g+f),$$

式中 g 为观测到的黑子群数,f 为观测到的单个黑子的总数,K 为换算因子,沃尔夫对自己的观测值取 $K=1$. K 值随观测者的观测技术、观测方法、使用的仪器和天气能见度等的不同而不同. 黑子相对数不仅能衡量太阳黑子活动的程度,而且也是衡量太阳活动的判据. 当黑子相对数增大时,太阳活动日趋活跃,在太阳活动峰年,黑子相对数达极大值. 黑子相对数的变化周期约为 11a.

沃尔夫相对数(Wolf relative number) 即"太阳黑子相对数".

太阳黑子周期(sunspot cycle) 太阳物理学的一个概念. 指太阳黑子活动变化规律所具有的周期性. 其周期为两个极小值之间的时间,长的约为 14a,短的约为 8.5a,平均长度为 11.1a. 开始 4a 中,黑子不断产生,活动加剧,以后 7a 中,大黑子群逐渐减弱、消失. 德国人施瓦布(Schwabe,S. H.)坚持 17a 对太阳黑子的观测,于 1843 年发现太阳黑子的消长有约 10a 的平均周期. 到了 1848 年,瑞士天文学家沃尔夫(Wolf,J. R.)整理了 1609 年天文望远镜发明以来的 200 多年太阳黑子记录,证实了太阳黑子活动周期的存在,周期平均长度为 11.1a,即太阳黑子 11a 周期. 黑子数从 1700 年起已有年总量值,1749 年起有更详尽的月总量值,故国际上统一规定,从 1755 年这一年黑子数最低点开始的那个 11a 周期作为第一个周期,往后顺次排出每个 11a 周期的号数,从 1986 年起开始进入第 22 个 11a 周期. 太阳黑子周期除 11a 基本周期外,1919 年,美国天文学家黑尔(Hale,G. E.)等人根据黑子磁场的极性分布经历一个循环约需 22a,提出太阳黑子 22a 磁性周期(亦称磁周期). 随着对太阳活动研究的深入,于 20 世纪中叶,格莱斯堡(Gleissberg,W.)等人又发现太阳黑子 80a 周期,此外,还有人提出许多更长的周期(178a、400a、1700a、2000a 的周期等)或更短的周期(小于 1a 的).

米粒组织(granulation) 太阳的一种物理现象. 指太阳光球层上的一种日面结构. 它在高分辨率的光球照片上,为暗条纹所分开而呈现米粒状的明亮斑点,故名. 米粒一般为不规则多边形,直径为 700~1800km,平均为 1000km(从地球上看约为 1.3″). 在同一瞬间,整个光球层约有 400 万个米粒,米粒间的暗条纹约为 290km 宽;米粒比暗条纹区域的温度高 300~400K,亮度约强 10%~20%;米粒组织变化很快,不断产生和消失,平均寿命约为 8min,个别可达 15min.

目前,认为米粒组织是光球层下面气体对流所形成的现象,米粒是上升的热气体柱的顶部,热气体柱把能量从内部带到光球层,而米粒间的暗条纹可能是下沉的冷物质. 于 1954 年还观测到超米粒组织,它是一种更大和更难观测的日面结构. 在整个可见的宁静太阳半球,约有 2500 个超米粒,直径为 $2\times10^4\sim6\times10^4$km,平均约为 3.2×10^4km,寿命为 20~40h,平均约为 24h. 据推测,超米粒的底部可能位于光球下面太阳对流层中 7000~10000km 的深处. 有人认为超米粒组织是由对流的不稳定性产生的;也有人认为超米粒组织的结构本身就是在对流层深处发生的磁场——对流运动相互作用的反映.

光斑(facula) 一种太阳光球层边缘的活动现象. 指在太阳光球层边缘出现的明亮斑点. 边缘上有黑子处就有光斑,但是没有黑子的地方有时也会出现光斑. 与黑子有关的光斑,呈纤维状,宽 5000~10000km,长约 50000km;寿命比黑子长,一般约长 3 倍. 与黑子无关的光斑,略呈圆形,面积较小,直径约 2300km,平均寿命约 0.5h,光斑温度比光球高,只是由于光斑不处于辐射平衡,底部温度要稍低一些,上层则高一些. 上层的平均温度比周围高约 100K,亮度大 10%左右. 光斑与黑子同样具有 11a 活动周期. 光斑向外延伸到色球层,便成为谱斑.

日浪(surge) 太阳的一种物理现象. 指太阳活动区内光球物质的急速抛射现象. 可与耀斑和小亮点共生. 它常在日面边缘,从一个小而亮的小丘顶部,以钉子形向外急速增长. 在 10~20min 内可达到 100~200km/s 的速度. 抛射最大高度为 $1\times10^4\sim2\times10^4$km. 抛射物质的总质量为 $10^{14}\sim10^{15}$g. 日浪爆炸区线度从几百千米到 5000km. 底部位于太阳黑子本影和半影内,或者在半影的外边界. 内部磁场强度为 0~0.015T. 具有轮廓分明的界限,经高分辨率观测分析,它是由一簇精细(小于 1″)的"纤维"组成,纤维间距约几(角)秒. 一般具有重复出现能力. 在抛射物质沿上升轨道返回太阳表面以后,又会触发新的日浪,但其极大速度和最大高度一次比一次小.

色球(chromosphere) 天体物理学术语. 指用肉眼无法看到的一层太阳大气层. 是太阳大气层中间的一层. 厚约 2000km. 色球层辐射的光很微弱,光球的光被地球大气散射,使太阳周围的天空很亮,完全淹没了色球,使平时单用肉眼无法看到色球. 日全食时由于月球遮住了光球圆面,天空变得昏暗,可

在黑暗的月球圆面周围看到一圈玫瑰色的光环，1869年，英国天文学家洛基尔(Lockyer,J. N.)和弗兰克兰德(Frankland,E. F.)将此光环命名为色球. 色球位于光球顶部和高温日冕之间，可分三层：低层，自光球层表面至约400km处，温度由4300K上升到6000K；中层400～1600km处，温度缓慢升到10000K；1600～2000km处为顶层，温度急剧上升到几万开. 色球气体稀薄，平均密度约10^{-9}kg/m^3. 色球是一个充满磁场的等离子体层. 由于磁场的不稳定性，常产生耀斑以及爆发日珥、日浪等现象. 还常有局部磁场较强的亮区(谱斑)出现.

针状物(spicule) 太阳的一种物理现象. 指从太阳色球层中层喷射到日冕层的明亮流焰. 它是宁静太阳上可观测到的最小景象之一，在边缘观测中看得最为清楚，像是直窜大气高层的火焰丛林，宽度平均为815km，视高度在极区约为9500km，在赤道附近约为7300km，真高度平均约为9800km. 数目随高度增加而递减，在日面3000km高度处约为93000个，在15000km处则只有约220个. 寿命平均约为5min. 以约25km/s的速度从色球层射出(有的匀速上升，有的跳跃上升)，使温度为10000K的色球物质伸展到温度更高的日冕层. 据认为，针状物把能量提供给日冕以维持其高温. 针状物的喷射速度小于太阳逃逸速度，因而绝大部分针状物在到达顶点后仍落回色球层，只有少数消失不见. 有些巨针状物则被喷射到达35000km的高度，而后也落回色球层上，这种巨针状物的能量是针状物的100倍.

谱斑(flocculus or plage) 一种太阳色球层中的活动现象. 用色球望远镜获取的太阳单色像上呈现的大块增亮区域，但有时也能观测到一些暗黑区域，因而将前者称为亮谱斑，后者称为暗谱斑. 谱斑位于色球层内、光球层的光斑上面，实际上谱斑就是光斑在色球层中的延续. 谱斑与光斑的形态相似，位置也基本一致；光斑只能在日面边缘观测到，而谱斑则在日面上大部分区域(包括日面边缘和日面中心)均可观测到. 谱斑与光斑存在有本质的物理联系，磁场便是这种联系的纽带. 谱斑与黑子也有密切的联系，谱斑常出现在大黑子和黑子群附近，黑子多时，谱斑也较多、较大和较亮，但寿命比黑子长(可持续几个太阳自转周期)，一般比黑子先出现而后消失. 谱斑的温度比周围高，大小从几千千米到几十万千米，形状、结构和亮度常在变化.

耀斑(solar flare) 亦称色球爆发. 一种太阳色球层中的活动现象. 指在太阳大气极小区域内发生的爆发性能量释放现象. 常出现在太阳黑子上空和附近. 一般只能用氢和钙单色光观测. 在有耀斑发生时，在色球望远镜中可看到局部区域在几秒到几分钟内突然增亮，在光谱中所对应的H_α谱线宽度和强度迅速增强，达到极大后慢慢下降. 一个大耀斑在短暂时间内(约10^2～10^3s)释放的总能量高达10^{23}～10^{26}J(接近太阳平时一秒钟辐射的总能量(3.9×10^{26}J)). 所辐射的除可见光外，有紫外线、X射线、γ射线、红外线和射电辐射，还有冲击波、高能粒子流和宇宙射线. 一般地，增亮面积超过3×10^9km^{2}的称为耀斑；不到3×10^9km^{2}的称为亚耀斑. 耀斑分为四级，分别以1,2,3,4表示，在级别后面加上f,n,b分别表示该耀斑亮度为弱、普通、强，所以最大、最亮的耀斑是4b，最小、最暗的是1f. 耀斑对地球影响很大. 爆发时，发出大量的高能质子流，严重危及宇宙飞船和人造卫星内的仪表与宇航员的生命安全；抛射出大量的带电粒子流，与地球高层大气作用时，常发生磁暴、极光和造成短波无线电通信的衰减甚至中断；对气象和水文等也有直接或间接影响.

色球爆发(chromospheric eruption) 即“耀斑”.

日冕(solar corona) 一层太阳大气层. 指太阳的最外层大气. 是一个极其稀薄、完全电离的等离子体区域. 主要由高度电离的原子和自由电子组成. 一般地，它延伸到几个太阳半径的地方，甚至到达地球轨道之外，在0.3太阳半径以内称为内冕；在0.3太阳半径之外称外冕. 日冕光线很弱，形状和亮度有11a的周期变化. 在黑子相对数极大年份呈圆形，光度大；在黑子相对数极小年份呈椭圆形，光度小. 气体温度为10^6K，电子密度随区域而异，在一个太阳半径以内平均约为10^8个/厘米^3.

冕洞(coronal hole) 日冕中的一个区域. 指日冕中一些辐射很弱、亮度比周围小得多的区域. 1950年，瑞士天文学家瓦尔德迈尔(Waldmeier,M.)首先从地面上观测到它，定名为冕洞. 分为三类：极区冕洞，位于两极区，常年都有；孤立冕洞，位于低纬区，面积较小；延伸冕洞，从北极区向南延伸至南纬约20°，或从南极区向北延伸至北纬约20°，且同极区冕洞相连结，面积较大. 寿命一般为5个太阳自转周，甚至1a. 温度约为10^6K，比宁静日冕区低. 密度约为宁静日冕区的三分之一.

冕洞中的磁场不均匀，孤立冕洞的磁场强度为1×10^{-5}～2×10^{-3}T；极区冕洞磁场强度约为0.0001T. 冕洞是太阳风的风源之一.

日冕凝聚区(coronal condensation) 日冕中的一个区域. 指日冕中明亮的区域. 1939年，由瑞士天文学家瓦尔德迈尔(Waldmeier,M.)发现. 是太阳表面局部活动区延伸到日冕后，日冕电子浓度比周围大的区域. 它的亮度和温度比周围高，寿命是几个月. 在不同波长上观测到的凝聚区有不同的名称，用光学方法观测到的称为光学凝聚区；用射电方法观测到的称为射电凝聚区；用软X射线观测到的称为

X 射线凝聚区. 经认证,上述三个凝聚区是同一个区域,但大小各有不同,日冕凝聚区的 X 射线辐射具有缓变成分,强度可以比周围区域大 70 倍.

冕流(coronal streamer)　日冕磁场不均匀分布的结果. 指日冕中比背景亮的延伸结构. 长度与太阳活动有关,在太阳活动极大时约延伸到 1 个太阳半径;在太阳活动极小时则可达 2 个太阳半径. 宽度在 0.1 个太阳半径以上. 可分为两类:盔状冕流(形状像钢盔);活动区状冕流(由活动区向外延伸,向外延伸的速度约为 2～10km/s).

日珥(prominence)　一种太阳活动现象. 它升腾在太阳表面的边缘. 其物质密度较大(比日冕大 10^3～10^4 倍),温度较低(约为 7000K),主要由氢组成,呈朱红色. 投影在太阳圆面上时日珥呈现为黑色纤维状,它比太阳圆面暗弱得多,称色球暗条,在一般情况下被日晕(即地球大气所散射的太阳光)淹没,肉眼无法直接观测,需用太阳分光仪、单色光观测镜等仪器,或在日全食时才能观测到. 可分为宁静日珥、活动日珥、爆发日珥、黑子日珥、龙卷日珥和冕珥等类型. 其形状则奇特多样,有的如浮云,有的如喷泉,还有的呈圆环、拱桥、火舌状等. 日珥的大小不一,一般长约 2×10^5km,高约为 3×10^4km,厚约为 5000km. 活动日珥和爆发日珥的上升速度可达 250km/s 以上,达到离日面 6×10^5～7×10^5km(有的达到百万千米)高度时再向色球层落下;有的日珥向上喷射后,长期悬浮在日冕中,既不坠落也不瓦解. 和黑子一样,日珥的多少同太阳活动强弱有关,有约 11a 的周期.

太阳活动(solar activity)　一种物理过程. 指发生在太阳大气局部区域和有限时间间隔中的物理过程. 主要表现为太阳黑子、光斑、谱斑、耀斑、日珥和太阳射电爆发等现象. 太阳活动的能源只占太阳总能源的极小部分,不涉及太阳本体的基本稳定性,只是太阳表层物质运动和变化所产生的结果. 太阳对流层的湍流场是调制太阳活动能源的枢纽,此外,太阳的自转和太阳对流层内的大尺度环流也是参与调制的因素. 由于它们的规律性和湍流场的随机运动,使得黑子、日珥、谱斑的出现和消失以及耀斑的爆发等太阳活动现象均呈现出既有规律性又有随机性的局面.

太阳活动区(solar active region)　亦称太阳活动中心. 太阳物理学研究的内容之一. 这里指在太阳大气层中太阳活动集中出现的区域. 太阳除了稳定地向日地空间不断发出均匀的辐射之外,有时在太阳大气层中也发生一些存在较短的"事件",例如,光球层出现的太阳黑子和光斑,色球层出现的日珥、谱斑和耀斑,日冕层出现的凝集区等. 这些现象并非均匀地分布在太阳大气层中,而是集中出现在太阳上的某些区域,这些区域便称为太阳活动区. 活动区的主要标志是太阳黑子群,在黑子群附近,几乎都有光斑,其上空的色球层中总是出现谱斑和经常有日珥,多数的耀斑也发生在太阳黑子附近. 同时,太阳黑子上空的日冕层中也常出现一些活动现象. 而且黑子群愈大,它附近出现的各种活动现象也愈多. 因此,人们通常用太阳黑子的多少来衡量太阳活动的强弱程度.

太阳磁场(solar magnetic field)　宇宙磁流体力学研究的内容之一. 指弥漫于太阳物质中的磁场. 在太阳表面的磁场可用光学方法观测到,在太阳内部的磁场是根据磁流体力学的定律和太阳黑子的状态推断出来的. 太阳外层的磁场可从日珥、耀斑、冕流和太阳风之间的关系中得出. 1908 年,美国天文学家黑尔(Hale,G. E.)首先根据太阳光谱线的塞曼效应测出太阳黑子的磁场. 现代采用不同类型的太阳磁场望远镜发现了太阳局部磁场、太阳整体磁场和磁结点等. 在太阳活动区内有局部磁场,在太阳黑子中,磁力线很集中,磁场强度可达 3×10^{-1}～4×10^{-1}T,1 个黑子群中有两个主要黑子,它们的极性相反;在太阳同一半球上,各黑子群的极性相同,在另一半球上的各黑子群极性刚好与此相反. 在一个太阳活动周期内(约 11a)结束,另一活动周期开始时,两半球的磁极性全部颠倒过来. 宁静日珥的磁场强度为 10^{-3}T,活动日珥的磁场强度可达 2×10^{-2}～3×10^{-2}T;除活动区外,太阳有一个微弱的普遍磁场,由于局部活动区磁场的干扰,仅在太阳两极区域比较显著,强度为 1×10^{-4}～2×10^{-4}T. 日面各磁场混合而成的整体磁场强度为 10^{-4}T,磁场强度有规律地变化,在一个太阳自转周期内极性由正变负,由负变正,极性变化两次. 近年来发现太阳磁场有很多复杂的精细结构,在某一磁极性的局部磁场中常出现呈另一磁极性的小磁结点. 此外,太阳风粒子将太阳磁场带到行星际空间,形成行星际磁场. 用理论方法估计,太阳内部磁场可能比太阳表面磁场强得多.

太阳磁场的起源还是个未解决的问题. 目前公认的一种学说是太阳平均磁流发电机理论,认为太阳磁场是由太阳对流层内磁流体的较差自转与湍流运动在磁场中所产生的感应电动势而维持的.

太阳振荡(solar oscillation)　天体动力学术语. 指太阳表面气体的一种周期为 5min 的起伏运动. 1960 年,莱顿(Leighton,R. B.)首先发现. 其特征是:

1. 在 10^3～5×10^4km 的范围内,同一高度的物质同时起落,振幅随到日面中心距离的增加而减少,随着高度的增加而增加,平均为 25km. 振荡是在垂直方向上进行的.

2. 在垂直方向上,不同高度处的振荡相位差是

不相同的. 5min 振荡可能与光球下面的对流层有密切关系. 由于对流层产生的波不能全部在光球大气里向上传播,有些频率的波只能在两层之间的区域内来回反射,形成不能传送能量的驻波. 这种重复的驻波可能形成 5min 振荡.

近年来观测表明,除 5min 振荡之外,还有其他周期的振荡. 1976 年苏联克里米亚天文台和英国伯明翰大学发现存在周期为 160min 的振荡. 美国天文学家狄克(Dicke,R. H.)发现太阳扁率有 12.64d 的周期振荡.

太阳自转(star rotation)　太阳物理学研究的内容之一. 指太阳本身的转动,与地球自转方向相同. 自转轴与黄道面法线成 7°15′的交角. 太阳为一高温等离子体球,其自转为较差式转动,即在日面不同纬度处,自转的角速度不同. 1859 年,英国天文学家卡灵顿(Carrington,R. C.)首先发现. 观测日面上的活动现象(如黑子、日珥、暗条、斑谱等)在日面上的移动,或者测定太阳东、西边缘光谱线的多普勒位移可以测定太阳表面的自转. 太阳自转角速度 Ω 可用下式表示:

$$\Omega=a+b\ \sin^2\theta+c\ \sin^4\theta,$$

式中 θ 为日面纬度,a,b,c 为常数. Ω 据日面活动现象的大量观测数据用最小二乘法归算得出,其值随活动现象不同而有区别. 通常采用 $\theta=17°$的值作为太阳自转周期. 以恒星为参考点,所得到的称为太阳自转恒星周期,为 25.38d;其角速度为 2.865×10^{-6}rad/s;相对地球观测者而得到的称为太阳自转会合周期,为 27.275d. 1904 年,哈姆(Halm,J. K. E.)首次注意到太阳自转速率是有变化的. 近代观测表明:太阳自转速率常有±10%的变化,变化规律和原因尚不清楚. 太阳色球和日冕的自转速率和光球相似. 在某些日面纬度上,日冕的自转速率比光球的要慢;而太阳磁场扇形结构的边界却呈现出刚性自转,而未呈现较差自转. 无法直接观测太阳内部自转,只能间接推测,不同学者得出不同结论,有人认为太阳内部自转速率比表面的要快,有的学者则认为内部自转速率比表面慢.

太阳常数(solar constant)　天文学术语. 指在地球大气外距离地球一个天文单位处,垂直于太阳光线的单位面积上每分钟所接收到的太阳总辐射能,通常以 S 表示,约为 8.25×10^3J/(m^2 · min)或 1.37W/m^2. 精确测定太阳常数比较困难,原因在于地球大气的吸收效应. 在地面(一般在空气稀薄的高山上)进行的测量精度为±1%,所达到的波段范围大约为 0.295～2.5μm,短于 0.295μm 和大于 2.5μm 的辐射被地球大气的臭氧、水气等吸收. 由于太阳辐射能量主要集中于可见光波段,因此,太阳常数所涉及的波段并不太宽,在 0.3～3.0μm 波段内的太阳辐射已占太阳常数的 97%. 20 世纪 60 年代以后,利用高空火箭、人造卫星、宇宙飞船测量太阳常数,便无需作大气吸收改正. 研究结果表明,太阳常数可能有小于 1%的变化.

太阳风(solar wind)　太阳物理学研究的内容之一. 指从太阳外层大气(日冕)不断发射出的稳定的粒子流,其中的粒子主要是质子和电子,其次是氦原子核. 日冕具有高达 10^6K 的温度,作用于日冕气体上的引力小于内外压力差,因此日冕并不处于静力学平衡状态,而是稳定地向外膨胀,这样的连续膨胀形成一种每秒带着约 100t 的气体并以每秒几百千米的速度向外流出的太阳风. 近年来的观测表明,冕洞是形成太阳风的风源.

太阳风的存在最早是从研究彗尾背向太阳的现象而推测出的,后来用空间探测器直接测量到太阳风. 长期观测表明,在地球轨道处太阳风的流速为 450km/s. 离子的数密度为 8×10^6 个/米^3,电子运动温度在 1×10^5～1.5×10^5K,磁场强度为 5×10^{-9}T. 太阳风的物理参数随太阳活动的强弱而有变化. 在太阳宁静期,太阳风的速度可降到 300km/s,密度为 3×10^6～6×10^6 个/米^3;随着太阳活动的增强,太阳风速可达 700km/s,密度达 3×10^7个/米^3;有时伴随大耀斑,日冕温度和密度会突然增高跃变到 4×10^6K 和 10^9 个/米^3,这时太阳风速可能达到 1000km/s.

太阳风的范围可延伸到 100 个天文单位而与星际气体混合在一起. 太阳风的高电导率意味着其中的磁场"冻结"在太阳风中,太阳磁场伴随膨胀的太阳风一道向外延伸而贯穿整个太阳系. 如果太阳没有自转,磁力线将是径向延伸,由于太阳自转,使磁力线扭曲成阿基米德螺线. 在地球轨道处,磁力线与径向成 45°交角.

太阳风及其磁场造成气体彗尾背离太阳,由于太阳风所携带的磁场向外扫掠,使太阳系内部区域宇宙线强度稍有下降,当太阳活动期,太阳风增强,宇宙线的强度下降到最低值. 太阳风将地球磁场压缩成一个彗星状的磁层.

太阳中微子(solar neutrino)　一种不带电的基本粒子. 指从太阳核反应区发射出的一种基本粒子. 静止质量接近于零. 中微子的存在是于 1930 年在研究原子核的 β 衰变时从理论上首先提出,于 1956 年在实验中被观察到. 中微子具有巨大的贯穿能力,因为它几乎没有质量,不会受任何物体的吸收;它又是一种不带电的中性微粒,电磁力也不能吸引它,能在任何一种原子核和电子之间任意穿行. 太阳内部氢核聚变反应过程中产生的大量中微子,能够不受阻碍径自射到地球,历时约 8min. 太阳中微子在地球表面处的通量很大,据理论预言,每秒钟发射到每平

方厘米地面的太阳中微子达600亿个，但是，经过多年观测，测到的中微子通量只有理论值的三分之一．两者相差悬殊，成为著名的太阳中微子之谜，至今仍未解决．

日地关系(solar-terrestrial relationship)　日地物理学的俗称．是太阳物理学和地球物理学之间的交叉学科．主要研究由太阳活动引起的太阳短波辐射和粒子流对地磁场、电离层的影响，同时也研究太阳辐射、太阳活动和气候变化之间的相互关系．太阳风、宇宙线中的带电粒子与地磁场的相互作用形成地球磁层及地球辐射带；太阳大耀斑、射电爆发发生时，发射出的高能带电粒子造成地球磁场的突然激烈变化——磁暴；太阳活动的增强会使地球电离层的离子浓度增高，而耀斑等激烈活动更会造成复杂的电离层效应：无线电短波衰减或中断（通常可持续几分钟到几小时），甚长波位相突然反常，长波增强，宇宙射电噪音突然吸收；大耀斑出现以后一两天内在地球高纬地区出现明亮的极光；此外，大耀斑发生时常发出大量质子及其他带电粒子，其能量可高达1000MeV，造成地面宇宙线增强，超过正常宇宙线三四个数量级；统计资料表明太阳黑子活动与地球降雨量相关十分明显．

太阳系物理学(solar system physics)　天体物理学的一个分支学科．研究太阳系的行星（行星大气、磁层和行星内部）、卫星、小行星、彗星、流星体和行星际物质等的物理特性、化学组成和行星际等离子体（太阳风）和磁场等．太阳系物理学按不同研究对象可细分为行星物理学、彗星物理学、流星天文学、陨星学和行星际空间物理学．20世纪60年代，由于空间天文技术的发展，空间探测器多次飞往月球，飞临水星、金星、火星、木星、土星、天王星、海王星等，使太阳系天体的研究从观测进入实验阶段，取得许多重大成果．

太阳系(solar system)　一种天体系统．指由太阳、九大行星及其卫星、小行星、彗星、流星体以及行星际介质组成的天体系统．在太阳系中，太阳是中心天体，它集中了太阳系总质量的99.86%，其他天体都在太阳引力作用下绕太阳运行．在太阳系中，九大行星的活动范围离太阳不到50天文单位，奥尔特云（参见“长周期彗星”）离太阳也只有10万天文单位的量级．太阳的引力作用范围还要大．太阳系只是庞大银河系的1千亿个成员之一，它离银河系中心1万秒差距（≈20亿天文单位），它绕银河系中心转一圈要2.5亿年之久．在太阳系中，九大行星的分布和运动的主要特征是：都在近于同一平面上的近圆形轨道，朝同一方向运动，它们与太阳的距离分布符合提丢斯-波得定则．九大行星可分成性质不同的三类：类地行星（水星、金星、地球、火星），其特点是体积、质量小，密度大；巨行星（木星、土星），其特点是体积和质量大，密度小；远日行星（天王星、海王星、冥王星），其特点是体积、质量小，密度大于巨行星，小于类地行星．除水星和金星外，行星均有卫星围绕，多数行星还有环带围绕．小行星绝大多数位于火星和木星轨道之间，各小行星的轨道倾角、偏心率彼此相差很大．

行星(planet)　一种天体．在偏心率不大的椭圆轨道上绕太阳公转的较大天体称为行星．太阳系共有九大行星，离太阳从近到远依次为水星、金星、地球、火星、木星、土星、天王星、海王星和冥王星．在地球轨道内的水星和金星称内行星，其余6个是外行星．根据其大小、质量、密度和化学组成可把行星分为类地行星（包括水星、金星、地球和火星）、巨行星（木星和土星）和远日行星（天王星、海王星和冥王星）三类（古典分成类地和类木两类）．行星主要靠反射太阳光而发亮．木星和土星等行星发出较强红外辐射，表明它们还有其他内部热源，这个热源不是热核能．现已发现除太阳外有多颗恒星拥有行星系统，如飞马座51、大熊座47、室女座70、仙女座υ、时钟座ιB、天鹅座16B、武仙座14和HD210277等．

行星视运动(planetary apparent motion)　一种天文现象．行星除和恒星、太阳等天体一起东升西落外，还在众恒星间移动，这种视运动称为行星相对于恒星的视运动，它也是行星和恒星名称的由来．这种视运动在黄道附近，当行星在恒星间由西往东运行时，称顺行，反之称逆行．顺行与逆行之过渡称为留．对于行星相对于太阳的视运动，内行星和外行星不同，因此，它们有着不同的出没规律．内行星只能徘徊在太阳附近，它经历：上合（内行星在比太阳远处与太阳黄经相等）→东大距（内行星在太阳东侧与太阳的角距最大）→下合（内行星在太阳和地球之间与太阳黄经相等）→西大距→上合．合附近，行星与太阳同升同落，看不见它；东大距附近，内行星傍晚出现在西方天空，它是昏星；西大距附近则是晨星，日出前在东方看到它．外行星相对于太阳的视运动经历：合（行星与太阳黄经相同）→西方照（黄经差90°）→冲（黄经差180°）→东方照→合．行星从合（上合）到下次合（上合）之时间间隔称1会合周期，它是行星出没规律重复的时间间隔．

提丢斯-波得定则(Titius-Bode law)　行星与太阳平均距离的经验规律．1766年，德国中学教师提丢斯(Titius, J. D.)指出：倘若令土星（当时是离太阳最远的行星）到太阳的距离为100，则水星到太阳的距离为4，金星为4+3=7，地球为4+6=10，火星为4+12=16，再向下，4+24=28的位置上未发现行星，过了这个空隙之后，木星为4+48=52，土星为4+96=100．1772年，柏林天文台台长波得

(Bode, J. E.)进一步研究了这个问题,使它广为流传,得名为提丢斯-波得定则. 1787 年,天文学家乌尔姆(Uim, H.)把它写成 $R_n=a+b\times2^n$ 的形式,若取天文单位为 1,则 $a=0.4$, $b=0.3$,水星 $n=-\infty$,金星 $n=0$,地球 $n=1$,火星 $n=2$,木星 $n=4$,土星 $n=5$,….

水星(Mercury) 太阳系的行星之一. 是离太阳最近的行星. 中国古代又称它为辰星. 水星轨道半长径为 5.79×10^7km. 它公转一周要 87.969d,自转一周要 58.646d,是公转周期的 2/3. 除冥王星外,水星的轨道倾角和偏心率最大. 水星近日点每百年进动 5601″,按牛顿力学只有 5558″,差 43″,这个反常进动被广义相对论圆满解释. 水星半径为 2440km,体积是地球的 0.0562 倍,质量是地球质量的 0.0558 倍,它的密度与地球差不多,这表明它有类似于地球的内核——铁核. 内核占水星总质量的 70%～80%,铁核外有 500～600km 的硅酸盐包层. 水星表面像月球,布满了大大小小的陨击坑. 水星大气很稀薄,昼夜温差特别大. 白天热到 700K,夜晚冷到 100K.

金星(Venus) 太阳系的行星之一. 是距太阳距离排为第二的大行星. 地球的近邻,最接近时离地球只有 4×10^7km. 从地球上看,它是最亮的行星,比最亮恒星天狼星亮 14 倍. 古代称它太白. 它有时黄昏后出现在西方,民间称长庚星;有时黎明前出现在东方,民间称启明星. 早在 1610 年,伽利略(Galilei, G.)就用望远镜看到金星有位相变化. 金星自转方向和其他行星自转方向相反,自转周期为 243d,比公转周期 224.7d 长. 金星的大小、质量和密度与地球差不多,因此它也有与地球差不多的核、幔层、壳层结构. 金星有浓密的二氧化碳大气层,温室效应使金星表面温度高达近 500℃.

地球(Earth) 太阳系的行星之一. 是距太阳距离排为第三的大行星. 是人类居住的星球. 它沿着偏心率 $e=0.0167$ 的椭圆轨道绕太阳公转,公转的轨道面称为黄道面,轨道半长径为 1.496×10^9km(=1 天文单位),公转周期为 365.256d(=1 恒星年). 地球自转一周时间为 23h56min4s(=1 恒星日). 地球赤道面与黄道面的夹角为 23°27′. 自转与公转的结合产生了昼夜和四季的交替,以及地面上的五个气候带. 地球形状近于旋转椭球体,赤道半径为 6378km. 地球内部有地核、地幔和地壳三层. 表面有陆地和海洋,海洋面积占 7/10,陆地则由高山和平原构成. 地球外部有大气层、磁层和高能粒子的辐射带.

地球辐射带(radiation belts of the earth) 地球外部的一种区域. 即地球磁场捕获带电粒子的区域. 它主要由高能电子和质子以及少量 α 粒子和重原子核所构成. 在地磁场作用下,带电粒子沿磁力线螺旋运动并不断辐射出电磁波,辐射带由此得名. 1905 年,斯托梅尔(Störmer, F.)预言辐射带的存在,1958 年,美国人范艾伦(Van Allen)发现了它,故辐射带也称范艾伦带. 辐射带分内辐射带和外辐射带两部分,其外形像套在地球外的两个环,子午截面呈月牙形. 内辐射带主要由高能质子组成,强度比较稳定;外辐射带主要由高能电子组成. 辐射带大小和外形不但与地磁场有关,而且与太阳活动亦有联系.

地球磁层(earth magnetosphere) 地球外部的一个包层. 指在太阳风场中地球磁场被禁锢的区域. 太阳风被排斥在地球磁层之外. 地球磁层内充满等离子体,它们的行为被地磁场支配. 地球磁层的外边界称磁层顶,磁层顶外是磁鞘,再外便是把太阳风速从超声突变到亚声的弓形激波面. 磁层顶内是等离子体幔,里面还有高能粒子辐射带等构造. 向日面磁层顶呈扁半球形,半径约为 10～12 倍地球半径;背日面有一圆柱形长长磁尾,其中间平面(赤道面上)磁场为 0,称中性片,片上下两侧磁场与片平行且方向相反. 向日面与背日面磁层由极尖区把它们分开. 太阳爆发形成之强太阳风会引起地球磁层的剧烈扰动,产生磁层暴和磁层亚暴.

黄道光(zodiacal light) 一种天文现象. 指在地球低纬度地区春季黄昏在西地平线或秋季黎明在东地平线上看到的淡弱的三角锥形光辉. 沿黄道向上伸展,直达地平高度 30°左右. 在离太阳 30°的地方,黄道光最亮,其亮度约是银河最亮处的 3 倍. 黄道光主要是由行星际尘埃散射太阳光而形成的,在黄道光中既有太阳的连续辐射,又有太阳的夫琅禾费吸收线. 黄道光中还有热辐射,太阳辐射被行星际尘埃吸收,再在红外波区向外辐射.

对日照(counterglow) 一种天文现象. 指在黄道带内与太阳相距 180°处的一个暗淡的弥漫状亮斑,大致呈卵形,范围约 20°×10°. 由于其亮度十分微弱,直到 1865 年,才为丹麦天文学家勃罗森(Brorsen, T.)首先观测到. 只有在远离城市的高山地区,在无月的晴夜才有可能看到对日照. 最有利的观测时间是 3 月和 9 月. 对日照的成因有多种假设,根据对日照光谱与太阳光谱极为相似这一事实,一般认为它是行星际物质反射太阳光而形成的.

火星(Mars) 太阳系的行星之一. 是距太阳系第四近的大行星. 中国古代把它称荧惑,肉眼看它发红色光辉. 火星沿着偏心率较大($e=0.093$)的椭圆轨道在黄道面附近绕太阳公转,公转周期约为 687d. 由于偏心率大,火星近日距与远日距相差甚大,达 4.2×10^7km,这就使火星冲日时与地球距离变化较大. 火星最接近地球的冲称大冲,它每隔

15～17a发生一次. 大冲时火星离地球只有 5.6×10^{11}km，是观测火星的良机. 火星自转一周约为24h37min，也有昼夜和四季交替. 火星比地球小，其赤道半径为 3395km. 也有核、幔和壳结构. 除两极地区外，其表面被沙漠所覆盖. 也有陨击坑、峡谷与河床(无水)及稀薄大气层.

火星尘暴(martain dust storm) 火星大气中的一种特有天气现象. 地区性尘暴在火星表面上经常发生. 全球性尘暴几乎每火星年发生一次，它往往发生在南半球春末火星过近日点时. 此时太阳光加热作用增强，引起气温的不稳定. 尘暴发源地位于阳光垂直照射的纬线上，往往在海腊斯盆地西边几百千米的诺阿哥斯地区. 强烈的全球性潮汐风与地形风相结合使地面尘埃扬起，尘粒被风带到空中，它吸收更多阳光而比周围更热，因而急速上升，四周空气去填补它，形成强地面风，使尘暴向四周扩展，几星期后覆盖南半球. 特大尘暴还能扩展到北半球，粉红-黄云遮盖全火星，这种现象可持续几个月之久.

火星极冠(cap of mars) 天文学术语. 指火星两极地区的白色覆盖物. 它是水冰和二氧化碳冰(干冰)层，其温度在－140～－70℃之间. 当火星北半球夏天时，北极冠缩小，南极冠扩大. 北半球冬天时则相反. 据估计，火星大气中有 1/5 的二氧化碳冻成干冰，冻结在两极冠中；而极冠中的水冰则比大气中的水要多得多. 如果把极冠中的水均匀分布在火星表面，则可形成一个 10m 厚的水层. 极冠不是整块的，有着分层结构.

火卫(satellites of mars) 火星的卫星. 即围绕火星运转的星球. 火星有两个卫星分别为火卫一(福波斯)和火卫二(德莫斯). 这两个卫星均于 1877 年火星大冲时，被天文学家霍尔(Hall，A.)所发现. 火卫一离火星平均距离只有 9000 多千米，绕火星一周只要 7h39min，比火星自转快，在火星上看，它是西升东落；火卫二离火星 2.4×10^4km，它绕火星一周要 30h16min. 两个火卫的自转周期等于它们的公转周期. 两个火卫形状都不规则. 火卫一大小为 27km×21km×19km，火卫二更小，只有 15km×12km×11km. 两个火卫上均有许多小陨击坑，其中最大的是火卫一上的斯蒂尼陨击坑，直径达 8km. 火卫一上还有沟纹及小环形山.

小行星(asteroid or minor planet) 太阳系中的一种天体. 它们大多数位于火星和木星的轨道之间，是沿椭圆轨道绕太阳公转的小天体. 1801 年元旦之夜，意大利西西里天文台台长皮亚齐(Piazzi，G.)发现第一颗小行星谷神星. 1802 年，德国医生天文爱好者奥伯斯(Olbers，H. W. M.)发现了第二颗小行星智神星. 1804 年，德国天文学家哈丁(Harding，C. L.)发现第三颗小行星婚神星. 1807 年，奥伯斯又发现第四颗小行星灶神星. 利用照相方法已发现许多小行星，其中算出轨道已获得永久编号的有 1 万多颗，而亮度亮于 21.2 等的小行星有 50 万颗之多. 最大的小行星谷神星直径约为 1000km. 直径大于 200km 的小行星约有 30 颗.

脱罗央群小行星(Trojan asteroids) 一类特殊的小行星群. 指绕太阳运行的轨道半径与绕木星运行的轨道半径相等的两个小行星群. 分别位于它们和太阳以及木星构成的平面圆形限制性三体问题的两个拉格朗日点 L_4 和 L_5 附近. L_4 和 L_5 是分别与太阳和木星构成等边三角形的两顶点. 这两群小行星中，在木星前 60°的一群为希腊群小行星，后 60°的为纯脱罗央群小行星. 从 1906 年发现第一个脱罗央群小行星阿基利斯开始，至今发现亮度亮于 21 等的脱罗央群小行星约有 900 颗，其中属希腊群的约 700 颗，属纯脱罗央群的约 200 颗. 它们一般均很暗，光谱特征也与其他小行星不同. 它们可能有特殊的起源.

小行星带(asteroidal belt) 围绕太阳的一个环形区域. 指位于火星和木星之间、轨道半长径 a＝2.17～3.64A(A 为天文单位)处的小行星组成的小行星环形区域. 常把匈牙利群(a＝1.9～2.0A)和希尔达群(a≈4A)小行星也包括在小行星带中. 小行星带内的小行星都顺向公转. 小行星带以外，就是些比较特殊的小行星. 小行星带中的小行星不是均匀分布的，有几条狭缝，很少有小行星能在其中留驻. 这几条狭缝称为扣克伍德环缝. 它们对应的轨道周期往往与木星的轨道周期成简单的整数比，称它们与木星轨道共振. 奇怪的是，有的共振区内没有小行星驻留，有的反而有密密麻麻的小行星群.

木星(Jupiter) 太阳系中的行星之一. 指太阳系最大的大行星. 中国古代又称它为岁星或太岁. 无论体积还是质量，木星均是九大行星之首，它的赤道半径为 71300km，为地球的 11.2 倍，体积为地球的 1316 倍. 木星公转周期为 11.86a，其自转，赤道快两极慢，赤道自转周期为 9h50.5min. 木星没有固体表面，是一个流体行星，它有内部热源. 在木星中心有一个主要由铁和硅构成的高温、高压、高密度固体核心，核心外是巨大的木星幔，幔由液态分子氢和金属氢组成，木星幔外便是浓密的大气. 木星幔物质对流为木星提供了强大的磁场和辐射带. 大气环流和飓风以及大气旋造成一系列与木星赤道平行的云带和斑块. 木星南半球著名的大红斑长 2×10^4km，宽 1.1×10^4km，它至少已存在 300a.

木卫(satellites of Jupiter) 木星的卫星. 指围绕木星运转的星球. 木星有 16 个卫星，其中 4 个大的卫星早在 1610 年就被伽利略(Galilei，G.)发现，有 3 个卫星是 1979 年旅行者 1 号和 2 号宇宙飞船

飞到木星时发现的.12个小卫星中,最大的木卫五半径为130km,其余卫星的半径只有几十千米乃至十几千米.木卫一是4个大卫星中最靠近木星的,其大小和质量均与月球差不多,它有强烈的火山活动,火山爆发放出大量的二氧化硫气体,它们构成了木卫一的大气层和电离层,并为木星磁层提供等离子体.木卫一表面平坦,有广阔的平原与起伏不平的山脉.木卫二、三、四的岩石表面上被厚厚冰壳层所覆盖,冰壳层上密布陨击坑.木卫三半径为2630km,是太阳系中最大的卫星.

伽利略卫星(Galilean satellites) 木星的卫星.指由伽利略(Galilei,G.)发现的4个卫星.木星的4个大卫星木卫一、木卫二、木卫三和木卫四,统称为伽利略卫星.1610年,意大利天文学家伽利略首先用望远镜发现,故名.伽利略卫星的发现为哥白尼(Kopernik,M.)的日心体系提供了重要的科学证据.

木卫一(Io) 木星的卫星之一.指最接近木星的伽利略卫星.1610年,意大利天文学家伽利略(Galilei, G.)发现.与木星的平均距离为412600km.按与木星接近次序排为第五.公转周期为1.769d,轨道偏心率为0.01,轨道倾角为0.03°.半径为1820km,质量为8.89×10^{22}kg,密度为3530kg/m^3,反照率为0.63,最亮时视星等为5.3等.有一固体表面,表面上有奇异的地貌:有弯曲的陡坡和断层,有众多的火山口,火山口周围为黑、黄、红、棕、橙色的辐射纹;有强烈的火山活动,已记录到9个活火山,火山爆发时以1km/s的速度喷射出高达70～300km的火山云;大气非常稀薄,表面气压约1×10^{-5}Pa,大气主要成分是二氧化硫;有一电离层,还有一个与木星相连结的等离子体环,表面温度为60～120K.

木卫二(Europa) 木星的卫星之一.指按其与木星的距离由近及远的次序排为第二的伽利略卫星.它与木星平均距离为671000km(按与木星接近次序排为第6).公转周期为3.551d,轨道倾角为0.46°,轨道偏心率为0.半径为1820km,质量为4.79×10^{22}kg,平均密度为3030kg/m^3,反照率为0.64,最亮时视星等为5.7等.其表面被冰层覆盖.赤道地带有斑状暗区和亮区及暗黑条带,条带宽数十千米,有的长达几千千米.冰层表面有巨大裂缝,最大的宽70km,深数千米,长1600km.一般认为条带为全球规模的断裂构造.

木卫三(Ganymede) 木星的卫星之一.指按其与木星的距离由近及远的次序排为第三的伽利略卫星.是木星卫星中最大的卫星,也是太阳系中最大的卫星,大于水星和冥王星,略小于火星.与木星平均距离为1071000km.公转周期为7.551d,轨道倾角为0.183°,轨道偏心率为0.002.半径为2640km,质量为1.48×10^{23}kg,平均密度为1930kg/m^3.反照率为0.43,最亮时视星等为4.9等.有一个很厚的冰壳表面,不规则的线状的明亮断裂带交叉穿越表面.表面上陨击坑密布的地貌与月球、火星、水星表面上陨击坑密布的高原非常相似.由古老的黑色大陨击坑组成的大面积圆形地貌约占可见半球面积的三分之一,一系列宽的明亮凹槽穿过其中,凹槽呈同心圆排列,稍弯曲而彼此平行.

木卫四(Callisto) 木星的卫星之一.指按其与木星的距离由近及远的次序排为第四的伽利略卫星.与木星的平均距离为1884000km.公转周期为16.689d,轨道倾角为0.253°,轨道偏心率为0.007.半径为2420km,质量为1.08×10^{23}kg,平均密度为1790kg/m^3.反照率为0.17,最亮时视星等为6.1等.有效温度为167K,是伽利略卫星中温度最高者.其表面密布陨击坑,也有冰壳层覆盖,与木卫三相似.

木星辐射带(Jupiter radiation belt) 木星外部的一种区域.指木星磁场捕获高能带电粒子的区域.它可分为质子带和电子带两部分.大于30MeV的高能质子带最大强度在3～4倍木星半径处,大于3MeV的高能电子带最大强度位于6倍木星半径处.高能质子主要靠磁层内部的加速.高能电子由于同步加速辐射失去能量而被靠近木星的木星卫星吸收.除了木星以外,土星、天王星和海王星均相继发现有辐射带.

土星(Saturn) 太阳系的行星之一.是太阳系的第二大行星.中国古代也称它为填星或镇星,其体积和质量仅次于木星.从望远镜中看土星,有一光环环绕它.土星自转和木星相似,也是赤道快两极慢,赤道自转一周为10h14min.赤道面与轨道面夹角为26°44′,故土星上也有昼夜和四季交替现象.土星和木星相似,主要由氢(占70%)和氦(占30%)所组成,它们有着相同的内部结构,只不过土星内部的压力、温度和密度均比木星小,土星幔的内层金属氢层比木星的深,土星核比木星的小.磁场和内部能源也比木星的弱.土星也有浓密大气层,大气中氢占94%.大气层中也有带纹、彩色云层、斑状结构,还有闪电与雷暴和强东西风.

土卫(satellites of Saturn) 土星的卫星.指围绕土星运转的星球.从1655年惠更斯(Huygens, C.)发现土卫六以来,现已确定土星有23颗卫星,它们构成太阳系中最大的卫星家族.在这23个卫星中,有10个由先驱者11号、旅行者1号和旅行者2号宇宙飞船发现,其余均由地面观测发现.土卫六半径为2575km,体积是月球的3.25倍,它是仅次于木卫三的第二大卫星.它有稠密的大气层、充满液体甲

烧的表面、冰幔和岩石核心，它的大气层厚度超过400km，比地球大气稠密．除土卫六外，其余卫星均是中等大小的或小的卫星，其中土卫八半球明亮、半球暗淡，两边亮度竟相差5～6倍之多．所有23个卫星中，平均密度小于$2000kg/m^3$，说明它们主要由冰组成．

土星环(Saturn rings)　土星外围的环形区域．指围绕土星运转的由小颗粒天体组成的环带．1610年，伽利略(Galilei，G.)首先看到土星球体旁有两个奇怪的附属物．1656年，荷兰物理学家惠更斯(Huygens，C.)认出这两个附属物是光环．1675年意大利天文学家卡西尼(Cassini，J. D.)在土星光环中发现一个空隙——卡西尼环缝．1856年，英国物理学家麦克斯韦(Maxwell，J. C.)从理论上证明了土星光环是由无数碎块构成的．土星光环包括3个主环和4个暗环．3个主环(从内到外)是C环、B环和A环．3个主环共宽约6×10^4km，其中B环既宽又亮，它宽2.5×10^4km，B环和A环间有5×10^3km宽的卡西尼环缝．旅行者号宇宙飞船发现土星光环是由成千上万条细而窄的环所构成，甚至在环缝中也能找到几条细环．

天王星(Uranus)　太阳系的行星之一．是距太阳距离排为第七的大行星．1781年3月，英国天文学家赫歇耳(Herschel，J. F. W.)用望远镜发现．天王星自转周期为17.24h，它的赤道面与轨道面交角为97.92°，它是“躺”在公转轨道上自转的．天王星的赤道半径为26 000km，是地球的4倍多，体积是地球的65倍，质量是地球的14.5倍，它的平均密度为$1240kg/m^3$．天王星表面被8000多千米深的汪洋大海所覆盖，海水温度高达几千摄氏度．海水上面的大气主要是氢(占85%～90%)和氦(占10%～15%)，其余气体不足1%．大气层中也有暗斑．天王星结构由岩-冰核、液态幔和富氢大气层三部分组成，有内部热源．天王星有磁场和辐射带，磁轴与自转轴交角为55°．

天卫(satellites of Uranus)　天王星的卫星．指围绕天王星运转的星球．天王星有15颗卫星，其中5颗是中等大小的卫星，半径在240～800km之间，其余10颗是半径只有几十千米的小卫星．5颗中等大小的卫星中，最大的天卫三和天卫四是1787年赫歇耳(Herschel，J. F. W.)发现的，天卫一和天卫二是1851年拉塞尔(Lassel，W.)发现的，1948年，凯珀(Kuiper，G. P.)才发现天卫五．其余10颗小卫星是旅行者2号宇宙飞船于1986年发现的．所有15颗卫星均在天王星赤道面中沿圆轨道绕天王星公转，它们都是规则卫星．天卫五表面有极其复杂的地貌．

海王星(Neptune)　太阳系的行星之一．是距太阳距离排为第八的大行星．发现天王星以后，计算它的位置总与观测不符，使人们联想天王星外还有一颗大行星，它的引力作用迫使天王星改变位置和运动．19世纪40年代中叶，英国的亚当斯(Adams，J. C.)和法国的勒威耶(Le Verrier，U. J. J.)各自计算了这颗未知行星的位置，柏林天文台加勒(Galle，J. G.)在预计位置附近找到它，取名海王星．海王星与天王星差不多大，也有核、幔和大气层的结构，大气层中也有卵形黑斑、暗斑和亮斑，也有磁场和环等结构．其磁场由冰幔中对流产生的发电机作用而产生，海王星被认为是磁场产生于行星核外的惟一行星．

海卫(satellites of Neptune)　海王星的卫星．指围绕海王星运转的星球．海王星有8颗卫星．1846年，德国的加勒(Galle，J. G.)观测到海王星后，同年英国的拉塞尔(Lassel，W.)就发现了海卫一，1949年，美国的凯珀(Kuiper，G. P.)发现海卫二，其余6颗均由旅行者2号宇宙飞船于1989年路经海王星时发现的．海卫一是太阳系中大卫星之一，直径为4000km，其体积和质量均比月球大，它在离海王星354 000km的圆轨道上绕海王星公转，它是逆行的，公转周期为5.877d．海卫一有稀薄的大气层，大气主要是氮气，大气层中还有一25km厚的雾云，它由冻结的甲烷所构成．海卫一上空8km处有火山云，证明它有火山活动．它有较大极冠，也有陨击坑结构．除海卫一外，其余7颗均是规则运行的小卫星．

行星环带(rings of planet)　一类行星周围的环形区域．指围绕行星运转的由小颗粒天体组成的环形区域．除了土星光环外，天王星、木星和海王星也有环．1977年3月10日天王星掩恒星时，美国、中国等五国天文学家观测到天王星有5条环，从内到外依次为α，β，γ，δ和ε环，其中最宽的ε环有着复杂的结构．后来又陆续发现6条环．1989年8月旅行者2号宇宙飞船发现海王星有5条环．外面两条是亮窄环，宽度不超过15km，另两条是暗而宽的弥漫环，还有一条是又宽又厚的尘埃壳层．最外的亮窄环物质分布极不均匀，环中有数十个亮点，可能是直径几千米至十几千米的小卫星．木星环是1979年3月旅行者1号宇宙飞船发现的，它由亮环、暗环和晕三部分构成．

冥王星(Pluto)　太阳系的行星之一．指离太阳最远也是最小的大行星．1930年2月18日，美国天文学家汤博(Tombaugh，C. W.)发现．冥王星绕太阳公转一周长达248a之久．它公转很特殊，不但轨道面大大偏离黄道面(倾角为17.1°)，其轨道偏心率($e=0.256$)也是九大行星之首，因此，虽然它轨道半长径近40A(A为天文单位)，它仍能走到比海王星离太阳还近之处．冥王星赤道直径约为2700km，

其体积只有地球的 0.009 倍，质量是地球的 0.0024 倍，平均密度约为 1500kg/m^3，与海王星和天王星的差不多. 日照面温度不超过 60K，背日面更冷，其表面绝大部分物质均凝固成固态或液态. 因此，冥王星上仅可能有极稀薄的大气.

冥王星卫星(satellite of Pluto) 冥王星的卫星. 指围绕冥王星运转的星球. 1978 年 6 月 22 日，美国海军天文台的克里斯蒂(Christy, J. W.)由冥王星星像扁长而推断冥王星有一颗卫星. 其后，智利托洛洛山天文台和美国麦克唐纳天文台均证实了这一发现. 此卫星取名为卡戎. 冥卫沿圆轨道绕冥王星运行，轨道半径为 1.9×10^4km，轨道面与冥王星赤道面重合，公转周期与冥王星自转周期一样，均是 6.387d，它是冥王星的天然同步卫星. 冥王星卫星的半径约 650km，不到冥王星半径的一半，其质量为冥王星的 1/10.

行星磁层(planetary magnetospheres) 行星周围的一种包层. 指行星固有磁场同太阳风相互作用，行星磁场的磁力线被禁锢的区域. 行星磁层内充满等离子体，其物理性质和过程受行星磁场的支配. 行星磁层的外边界在向日方向是清晰的，行星磁场的磁力线与外面太阳风的磁力线之间有明显的界线，其界面位于太阳风的动压与行星磁场磁压相等的地方. 在背日面，行星磁场的磁力线与太阳风中的磁力线混在一起，没有明确的界面. 在太阳系九大行星中，水星、地球、木星、土星、天王星和海王星均有磁层. 水星磁层很小，结构类似于地球磁层. 木星磁层很大，结构很复杂.

卫星(satellite) 一种天体. 指围绕行星公转的天体. 至目前为止，在太阳系九大行星中共发现有 66 个卫星，其中月球是地球的惟一天然卫星，火星有 2 个，木星有 16 个，土星有 23 个，天王星有 15 个，海王星有 8 个，冥王星有 1 个. 在这 66 个卫星中，比月球大或与月球差不多大的有 7 个，它们的半径超过 1500km，其中木卫三最大，半径为 2900km，它的体积是月球的 3.5 倍，比水星还大. 有 19 个是中等大小的卫星，其半径在 100～800km 之间. 其余的 40 个是半径只有几十千米甚至更小的小卫星.

月球(Moon) 亦称月亮. 一个天体. 是地球惟一的天然卫星. 它绕地球公转，同时又跟着地球绕太阳运转. 月地平均距离为 3.84×10^5km. 月球公转轨道在天球上的投影是一大圆，称白道，它与黄道的交角平均为 5°9′. 月球自转周期与它的公转周期相等，都是 27.32d. 所以，月球总是以同一面(正面)朝着地球. 月球赤道半径为 1738km，它的体积是地球的 1/49，质量是地球的 1/81. 月球几乎没有空气. 月面上有陆地、山脉和峭壁、月海、月坑以及月面辐射纹等结构. 月球可分为月壳、月幔和月核三层.

月龄(Moon age) 天体物理学的一个概念. 指从朔起算，到各月相所经历的时间. 月相和月龄的关系平均为：上弦月龄 7.4d、望 14.8d、下弦 22.1d. 朔日在农历每月初一，加上对应月龄，得到对应月相呈现的时间：上弦发生在农历初八，下弦在二十三，望在十五或十六. 由于月球运行速度不均匀，上面的时间可能有一天的差异.

月相(phase of the Moon) 一种天文现象. 指月亮盈亏圆缺变化形成的现象. 月球本身不发亮，它是被太阳光照亮的. 月球绕地球公转时，被太阳光照亮的半球有时朝向地球，有时背向地球，形成有时圆有时缺的月相变化. 当月球和太阳的地心黄经差为 0°，90°，180°和 270°时，对应的月相分别为朔、上弦、望(满月)和下弦. 月相更替的周期平均为 29.53d，称为一个朔望月.

月坑(lunar craters) 亦称环形山. 月球表面的一种地形. 指月球表面上的圆形凹坑. 月坑大小不一，小的直径只有几十厘米或更小，大的超过 200km. 最大的月坑是贝利月坑，直径为 295km，克拉维月坑直径为 240km，牛顿月坑直径为 230km. 直径超过 1km 的月坑有 3 万 3 千个之多. 月坑形态和构造多种多样：小的月坑只是一个浅的凹坑；大的月坑周围往往有环形山，它一般高出月面 300～7000m，月坑中央往往还有山峰或峰群；有的月坑周围有辐射纹. 月坑底部一般是平坦的，深度从几米到几千米. 绝大多数月坑是小天体陨击形成的，可能也有少数是火山坑.

环形山(craters) 即"月坑".

月海(maria) 月球表面的一种地形. 指肉眼看到的月面上较暗的区域. 月海表层没有水，它是月面上广阔的平原地区，为玄武岩所覆盖. 大的月海称为洋，小的称为湖，伸向月陆(高出月海的地区)的称为湾. 全月面共有 22 个月海，大多数在正面，仅有 3 个在背面. 在正面，月海面积占一半. 最大月海是正面的风暴洋，面积约 5×10^6km^2，有半个中国大，是月球总面积的 1/30. 其次是雨海，面积约 9×10^5km^2. 还有静海、澄海、丰富海、危海、云海等. 月球背面还有些直径约为 500km 的圆形凹地，它们未被熔岩物质充填，称为类月海.

月面辐射纹(bright rays on lunar surface) 一种天文现象. 指从某些年轻月坑向四周呈放射状延伸的亮条纹. 典型的有第谷月坑辐射纹，共有 12 条辐射纹，最长的达 1800km，满月时用望远镜能清楚地看到它们. 此外，哥白尼月坑的辐射纹也很清楚. 现在认为，辐射纹是小天体陨击月面时从月坑中抛出的物质堆积而成，也有人认为是月球火山灰四溅而形成的.

反照率(albedo) 用来表示不发光天体反射光

线本领的物理量.天文学中常用几何反照率、邦德反照率等多种反照率.其中邦德反照率(即球面反照率)最常用,它表示天体表面向各方向反射之总光流与入射光流之比.暗黑物体比白色物体反照率低.反照率为1的天体是纯白的,它能把入射到它表面的光线全部反射掉;反照率为0的天体则是纯黑的,它把入射光线全部吸收.行星和卫星的反照率大小,表明覆盖其表面上的物质特性.

彗星(comet) 俗称扫帚星.太阳系中的一类小天体.指太阳系中围绕太阳运转、靠反射太阳光和气体荧光作用而发光的一种小天体.通常,彗星在偏心率很大的椭圆轨道乃至抛物线或双曲线轨道上绕日运行.当它远离太阳时,仅仅是一个孤零零的彗核——一个几千米至几十千米大小的冰和尘埃冻结的"脏雪球".使用最大望远镜在地面上也难看到它.当它走近太阳时,由于阳光加热,冰雪升华为气体,并带出尘埃和冰粒,从而形成彗发——彗星的大气层.用望远镜看上去,彗星是一个大体呈球形的雾状亮区.当它离太阳足够近时,从彗发向反太阳方向上延伸出长长的彗尾(个别彗星有向日彗尾).此时,它发育为一个有彗头(彗核和彗发的总称)和彗尾的完整彗星.通常彗尾有两支:短、弯、粗的是尘埃彗尾(太阳辐射压把彗发中的尘埃推离彗星的尘埃粒子流);长、直、窄的是离子彗尾(太阳风吹走彗发中的离子和电子的带电粒子流).空间观测还在彗发外发现有彗晕,它是由极其稀薄的氢原子云和羟云所构成.通常,彗发直径可达几万乃至几十万千米,尘埃彗尾长几百万至上千万千米,离子彗尾长几千万至上亿千米,彗晕直径可达几百万乃至上千万千米.由于太阳风和彗发中气体相互作用,不但使彗头中有着复杂的结构,也使离子彗尾千变万化.

短周期彗星(short-period comet) 一种彗星.指绕太阳运行一周短于200a的彗星.到目前为止,已发现约100个短周期彗星.恩克彗星是一颗周期最短的彗星,它的周期只有3.3a,因而它出现次数最多.自1786年1月17日首次发现它以后,至今已观测到它60多次.在这100个短周期彗星中,约有70个彗星远日点在木星轨道附近,称它们为木星族彗星.

长周期彗星(long-period comet) 一种彗星.指绕太阳公转周期大于200a的彗星.已发现长周期彗星总数约为530个,其中有30%沿很扁的吻切椭圆轨道运动,54%沿抛物线轨道运动,16%沿双曲线轨道运动.轨道最大偏心率为1.006.统计研究表明,长周期彗星来自太阳系的边远地区.在离太阳3万到10万天文单位处,约有1千亿个彗星.这个巨大的彗星仓库称为奥尔特云.受太阳附近的恒星的引力摄动,奥尔特云中少数彗星轨道改变,有的进入太阳系内,成为人们看到的"新"彗星.

哈雷彗星(Halley comet) 最著名的一颗彗星.它在1682年出现时,英国天文学家哈雷(Halley,E.)注意到它的轨道和1607年以及1531年出现的彗星轨道差不多,认为它们是同一彗星的三次出现.哈雷于1704年精确算出了这颗彗星的轨道,并预言它在1758年底或1759年初将再度出现.1759年3月,这颗彗星果然再度回归.为纪念哈雷的功绩,把它称为哈雷彗星.1986年哈雷彗星回归时,有6艘宇宙飞船去考察它,发现其彗核为三轴长分别为15km、8km、8km的椭球,其表面高低不平,有几个直径为1km的浅坑及丘、谷,表面覆盖一层不均匀暗尘.中国史书上有世界公认的最早和最全的哈雷彗星出现记录.史书《春秋》记载:鲁文公十四年(公元前613年),"秋七月有星孛入北斗".这是最早的哈雷彗星出现记录.

流星体(meteoroid) 太阳系中一类小天体.指太阳系中运行于行星际空间中的碎小天体.它们各自沿一定轨道绕太阳运行.当流星体与地球相遇时,往往以11~72km/s的速度冲向地球,与空气分子激烈碰撞而灼热,在大气中燃烧而形成一道亮线,此即人们看到的流星.肉眼看到的流星,亮度一般要亮于5等.晴朗无月之夜,肉眼平均每小时看到10个流星.其中大多数亮度在1~4等之间,以2.5等最多.若用望远镜或雷达观测,则能看到更多更暗的流星.流星体多数在120km高空开始燃烧,到80km高空燃烧完毕,化为灰烬而消失.

火流星(bolide or fireball) 一种天文现象.指较大的流星体在地球高层大气中未燃烧完而进入地球低层大气中燃烧而产生的发光现象.此时不但看到流星特别明亮,有的比满月还亮,个别的照亮大地如同白昼,在流星经过的路上还会听到沙沙声响.火流星熄灭后几分钟内,也会听到噼噼啪啪的爆炸声和隆隆雷声.

流星群(meteor stream) 太阳系中一类小天体.指沿同样轨道绕太阳运行的众多流星体.有的流星群是由彗核释放出的尘粒所形成.1866年证实,英仙座流星群与彗星1862Ⅲ有相同的轨道,表明英仙座流星群是彗核破裂的产物.如果彗核碎裂而瓦解的话,则构成更大的流星群,仙女座流星群就是比拉彗星彗核碎裂瓦解的产物.比拉彗星于1865年失踪,分裂粉碎为仙女座流星群.

流星雨(meteor shower) 一种天文现象.指在同一天区每小时能看到十几个到几十个流星的现象.地球与流星群相遇时,就会看到流星雨现象.它们好像由天空同一点向外辐射,这一点称为此流星雨的辐射点,一般用它所在的星座来命名此流星雨.每年8月中旬,地球穿过英仙座流星群,人们会看到

英仙座流星雨. 每年 11 月中旬，可看到狮子座流星雨. 每隔 33a 左右可观测到一次狮子座流星雨暴：1799 年，委内瑞拉等地看到了成千上万个流星落了 4 个小时；1833 年，美国看到流星像雪花那样多，估计有 24 万个流星降落；1782 年 11 月 27 日，欧洲观测到英仙座流星雨暴，从下午 7 时到凌晨 1 时，有 16 万个流星降落.

流星余迹(meteor trail) 一种天文现象. 指流星体以高速与地球大气分子碰撞，使空气分子灼热、激发和电离形成的电离气体柱. 其直径约 1m 左右，长几十千米. 可用流星余迹反射无线电波来实现地面上不同地点的无线电通信. 一般亮于 12 等的流星都可用来作为远距离无线电通信的媒介.

陨星(meteorite) 亦称陨石. 天体碎块. 指从天外穿过地球大气层时未燃烧尽而落到地面上的天体碎块. 目前，全世界已收集约 3000 颗陨石，根据其化学成分，分为三种类型：石陨石、铁陨石和石铁陨石. 其中石陨石最多，占总数的 93%，铁陨石占 6%，其余 1%～2%是石铁陨石. 在南极找到 4 块陨石，经研究，认为它们是月球的岩石，其中 1 块是月球火山岩，其余 3 块是月球高地的岩石. 全世界收集到 8 块可能来自火星的陨石，它们可能是火星表面受小天体陨击而飞到太空的岩石，它们保留了火星表面的许多特征.

石陨石(stony meteorite) 一种陨石. 指主要由铁-镁硅酸盐和金属铁镍所组成的陨石，是陨石中数目最多的一类. 其金属铁镍含量一般小于 25%. 按其结构又分为球粒陨石和无球粒陨石两种，其中球粒陨石最多，占 84%. 球粒陨石内部一般都散布着许多直径为零点几毫米至几毫米的球状颗粒，球粒的主要成分是橄榄石和辉石. 全世界最大的石陨石是吉林陨石雨中的 1 号陨石，重 1770kg，它是含铁量较高的球粒陨石，铁镍合金及陨硫铁占 25%. 20 世纪 70 年代，美国科学家分析了两块碳质球粒陨石，首次发现陨石中存在有机化合物.

铁陨石(iron meteorite) 一种陨石. 指主要由铁、镍所组成的陨石，其特征是镍含量高. 地球上自然铁中镍含量不超过 3%，一般在 1%以下，而铁陨石中镍含量往往超过 5%. 全世界最大的铁陨石是非洲的戈巴陨铁，重约 60t. 中国的新疆大陨铁重 30t，是世界上第三大陨铁.

玻璃陨石(tektite) 一种陨石. 这种陨石是一种天然的玻璃体，不透明，颜色以黑色和棕黑色为多. 一般只有几厘米大小，有液滴、哑铃、钮扣等各种形状. 现代研究表明，它是小天体猛烈撞击地球表面岩石飞溅出来的熔融物质冷却而成. 地球上玻璃陨石集中分布于亚澳、捷克、北美和象牙海岸等四个散落区. 中国是世界上最早记录有玻璃陨石的国家. 中国古代把玻璃陨石称雷公墨，早在一千多年前，唐朝书中就有关于雷公墨的记载. 中国玻璃陨石主要分布在雷州半岛和海南岛一带，属亚澳散落区. 亚澳散落区可能是 70 万年前同一次陨击事件造成的.

陨石雨(meteorite shower) 一种大规模的陨星陨落现象. 是大流星体在天空中爆烈成碎块，而同时进入地球大气形成的. 1976 年 3 月，降落在中国吉林地区的吉林陨石雨是石陨石雨，现共收集到 100 多块陨石，总重约 2500kg. 对吉林陨石雨研究表明，它是在 47 亿年前从太阳系原始星云中分离出来的，这个年龄大体与地球形成年龄差不多. 这团云逐渐冷却与凝聚，形成许多液滴和结晶体，它们相互碰撞、破裂与聚合，形成许多小团块，相互吸积使小团块合并构成一个半径为 200km 左右的小行星. 受其他小天体碰撞，大约在 800 万年前，这颗小行星被碰出一块巨大的流星体，它就是吉林陨石雨的前身.

行星际介质(interplanetary medium) 一种星际物质. 指行星际空间中的稀薄气体和尘埃. 黄道光、对日照以及 F 日冕均是行星际尘埃存在的证据，它们主要是由行星际尘埃散射太阳光而形成的. 行星际尘埃大小在 5～100μm 之间，其主要化学成分是镁-铁硅酸盐、钙长石、钙钛矿以及铁钙金属等. 彗星和小行星的瓦解是行星际尘埃的主要源泉. 太阳风是来自太阳的高速电子和质子流，它是行星际气体的主要源泉.

太阳系起源(origin of the solar system) 天体演化学研究的内容之一. 它研究太阳系总体和其各类成员的形成以及现今太阳系天体各种特征的由来. 太阳是颗恒星，它的起源和演化虽与行星系的形成有密切的关系，但一般在恒星的起源与演化中研究. 太阳系起源的第一个基本问题是行星物质的来源. 这方面的学说基本上可分为三类：灾变说(或分出说)、俘获说和共同形成说. 灾变说认为行星物质是因某一偶然事件(例如，另一恒星走近或碰到太阳)从太阳中分出来的；俘获说认为太阳从恒星际空间俘获物质，形成星云，星云后来演变成行星；共同形成说认为整个太阳系所有天体都是由同一原始星云所形成，星云中心部分物质形成太阳，外围物质形成行星和卫星等天体. 后两类学说的共同点是认为由星云凝聚成行星，常合称为星云说.

太阳系起源的第二个基本问题是行星的形成方式问题，大致有五种看法：先形成环，由环再形成行星；先形成很大的原行星，原行星再演化成行星；先形成中介天体，由中介天体结合成行星；先形成湍涡流的规则排列，在次级涡流中形成行星；先凝聚成大大小小的固体星子，星子再集聚成行星. 1644 年，法国哲学家和数学家笛卡儿(Descartes，R.)在他的《哲学原理》一书中提出了旋涡学说，从而开始了太

阳系起源的研究. 笛卡儿认为，在太初的混沌中，物质微粒获得旋涡运动，在旋涡中形成太阳、地球、行星和卫星. 1745 年，法国动物学家布丰(Buffon，G.-L.L.)首先提出了灾变说. 而真正把太阳系起源学说建立在比较科学的基础上的是 18 世纪的康德(Kant，I.)和拉普拉斯(Laplace，P.-S.)各自提出的星云说. 这两个学说风行世界达一个多世纪之久. 由于康德和拉普拉斯星云说无法解释太阳系角动量的特殊分布，19 世纪末至 20 世纪 40 年代初，各种灾变说一度盛行起来. 灾变说人为地假定许多偶然因素起主导作用，特别强调一颗恒星走近或碰到太阳的作用，然而银河系中两恒星相遇的概率非常非常小，几乎是不可能的，于是灾变说又逐渐失势，人们又回到星云说. 1944 年，苏联地球物理学家施米特(Шмидт，О.Ю.)提出俘获说；同年，德国物理学家魏札克(Weizsäcker，C.F.V.)提出旋涡说；1949 年，美国天文学家凯珀(Kuiper，G.P.)提出原行星说；瑞典物理学家阿尔文(Alfvén，H.)的电磁说则强调电磁和等离子体过程的作用；1952 年，美国化学家尤里(Urey，H.C.)提出的学说重视化学过程的作用；苏联的萨弗罗诺夫(Сафронов，В.С.)、日本的林忠四郎等人的星云说都较详细地论述了行星的形成问题，目前较受重视；美籍天文学家黄授书 1965 年提出了他对太阳系早期历史的看法；中国天文学家戴文赛对太阳系起源也提出一种见解. 目前，共有 40 多家太阳系起源学说，尚没有一家被世人所公认.

太阳系起源的研究至今已取得不少成果. 由太阳系的同位素丰度得知，太阳系原始星云约在 50～58 亿年前形成. 根据对恒星的起源和演化推断，太阳是在约 50 亿年前由星际云瓦解后的一团小云坍缩形成的，它经历了约 4 千万年的引力收缩阶段. 从地球和月球的古老岩石和陨石的放射性分析得知，地球和月球约在 46 亿年前形成，因此，太阳系应在 46～50 亿年前形成. 大行星形成后，发生过地球史所经历过的那种地质变化，因此，大行星现况与形成时不同. 小天体形成后质变较少，保留了较多形成时的信息，因此，研究彗星、陨石等有助于了解太阳系的早期情况. 月球、水星、火星等天体表面大多数凹坑，是 39 亿年前小天体陨击而形成的，这一事实对星云说特别有利.

星云说(nebular hypothesis)　一类太阳系起源学说. 该学说主张太阳系由一个星云所形成，或认为行星和卫星是由围绕太阳的星云盘所形成. 笛卡儿(Descartes，R.)于 1644 年在他所著的《哲学原理》一书中提出的旋涡学说是第一个星云说. 康德(Kant，I.)和拉普拉斯(Laplace，P.-S.)的星云说从数学和力学上作了论证，为研究太阳系起源奠定了基础，从而在 19 世纪被人们普遍接受. 由于科学的发展，人们发现这两个学说中也有不少缺点和错误，20 世纪又先后提出了 20 多种现代星云说，它们都较全面地论述了太阳系的起源，都对太阳系的各种特征或一部分特征作了说明.

灾变说(catastrophic hypothesis)　一类太阳系起源学说，该学说主张太阳系的形成是宇宙中某种偶然事件灾变的结果. 1745 年，法国动物学家布丰(Buffon，G.-L.L.)提出第一个灾变说，认为一个巨大彗星掠碰到固态太阳的边缘，它不但推动太阳的自转，还碰出一些物质绕太阳旋转，这些物质最后形成行星. 但彗星质量比地球质量小得多，绝对不会碰出这么多物质来形成行星和卫星. 后来，又陆续提出了十几种灾变说，试图解决现今太阳系角动量的特殊分布. 这些灾变说都认为是一个巨大的恒星走近或碰到太阳，从太阳分出物质形成行星. 这些学说中以英国天文学家金斯(Jeans，J.H.)的潮汐说最著名.

地心体系(geocentric system)　一种宇宙构造学说. 这一学说最初由亚里士多德(Aristotle)提倡，公元 130 年左右，被天文学家托勒密(Ptolemy)所发展，形成了托勒密地心体系. 其主要论点是：地球是静止不动的，位于宇宙中心；太阳、月球和行星都绕地球做匀速转动；恒星都位于"恒星天"的固体壳层上，日、月和行星除绕地球匀速转动外，还与"恒星天"一起，每天绕地球转一周，于是各种天体每天均要东升西落一次. 地心体系能大致解释行星的视运动，编制行星星历表，起过进步作用. 后来被教会利用，严重阻碍了科学的发展.

本轮(epicycle)　天体物理学术语. 指托勒密地心体系中引入的数学模型. 为解释行星视运动不均匀性，例如，出现顺行、逆行、留等现象，又保持匀速圆周运动不变，托勒密(Ptolemy)的地心体系中采用了古希腊学者阿波罗尼奥斯(Apollonius，(P))提出的本轮和均轮之偏心模型. 托勒密认为行星在一个称为本轮的小圆轨道上匀速转动，本轮的中心在称为均轮的大圆轨道上绕地球匀速转动，地球不在均轮的圆心上，而是偏离一段距离. 随着航海事业的发展，对行星位置观测精度越来越高，一个本轮已无法解释观测结果. 后来人们只好在本轮上又加上一个又一个更小的本轮，构成了本轮系，但也无法十分精确地计算行星位置，终于被日心说所取代.

均轮(deferent)　见"本轮".

日心体系(heliocentric system)　一种宇宙构造学说. 是由波兰天文学家哥白尼(Kopernik，M.)创建的. 其主要内容有：水星、金星、火星、木星和土星五颗行星与地球一样都在圆轨道上绕太阳公转；月球是地球的卫星，它在圆轨道上绕地球公转；天穹

不动，恒星非常遥远；地球每天自转一周，从而引起天体东升西落现象.

哥白尼日心学说的提出，打破了托勒密地心说对科学发展的桎梏，为建立科学的太阳系结构打下了基础，使自然科学从中世纪的神学束缚中解放出来.在天文学史上称为天文学上的革命.

康德星云说(Kant nebular hypothesis) 一种太阳系演化假说.1755 年，康德(Kant，I.)提出一种星云学说，认为太阳系是由一团微粒通过万有引力而形成的.引力使微粒相互接近，大微粒把小微粒吸引过去凝成较大的团块，团块越来越大.引力最强的中心部分吸引的物质最多，先形成太阳.外面的微粒在太阳吸引下向中心下落并形成几个引力中心，它们分别凝聚成行星.1796 年，拉普拉斯星云说问世.这一学说认为形成太阳系的云是一团巨大的、灼热的、转动着的气体，由于冷却，星云收缩，转动加快，星云变为盘状.当离心力和引力相等时就有部分物质留在原处，演化为一个绕中心体转动的环，以后陆续形成几个环.星云中心部分凝聚成太阳，各环凝聚成行星.

拉普拉斯星云说(Laplace nebular hypothesis) 见“康德星云说”.

戴文赛太阳系起源学说(Dai Wensai hypothesis for origin of the solar system) 太阳系演化假说.是中国天文学家戴文赛提出的.这一假说认为：康德和拉普拉斯星云说的基本论点是正确的，整个太阳系是由同一原始星云所形成.戴文赛认为原始星云是从星际云中瓦解出来的，一开始就有自转，并靠自吸引而收缩.星云中心部分形成太阳，外部因自转而扁化为星云盘，在盘中形成行星和卫星.原始星云既不是康德星云说中的微粒，也不是拉普拉斯星云说中的高温气体云，而是低温气体尘埃云.关于行星的形成方式问题，他认为星子集聚之论点是合理的.星云盘内固体颗粒沉降到赤道面，形成尘层，尘层瓦解形成星子，星子集聚成行星.

恒星物理学(stellar physics) 天体物理学的一个分支学科.它应用天体测光和天体光谱学等天体物理方法研究各种类型恒星的形态、结构、物理状态、化学组成、内部结构以及恒星的诞生和演化.使用光学、红外、射电和 X 射线等各种望远镜和检测手段(照相、光电、分光、偏振、测热、微波、能谱等)，接收恒星的电磁辐射、高能粒子和引力波等，测量各类恒星的辐射强度、角直径、角距离、角位移、视面结构、能谱、谱线结构、偏振等.然后运用物理学理论推断恒星的各种不同特性，如表面有效温度、恒星大气电子压力、气体压力、不透明度、元素的丰度、恒星的光度、恒星自转以及双星轨道半长径、子星半径、子星质量、脉动变星的平均半径和平均密度等.综合应用各种物理学理论，可以探测恒星的形成和演化规律.目前，由于全波段观测手段的广泛应用，加上现代物理学的发展、大型电子计算机的应用，恒星物理学研究已进入新的发展阶段.

恒星(stars) 一类天体.指炽热等离子体组成的自身发光的巨大星球.太阳是无数恒星中很普通的一颗.除太阳外，恒星离地球都十分遥远，最近的恒星距地球 4.2 光年.恒星虽然都在运动，但由于距离遥远，人们看来，它们在天球上每年移动角度极小，所以称之为恒星.

恒星质量与太阳相差不大(最大的为太阳的 60 倍，最小的为太阳的 1/30)，但大小和亮度却相差很大(直径最大的比太阳直径大千倍以上，小的是太阳的几十万分之一；最亮的比太阳亮几十万倍，暗的比太阳暗几十万倍).恒星表面温度从 2000K 到几万开.中子星的表面温度高达 10^7K.恒星中心温度可达 2×10^7K 甚至上亿开.维持恒星辐射的能源主要是原子核反应.

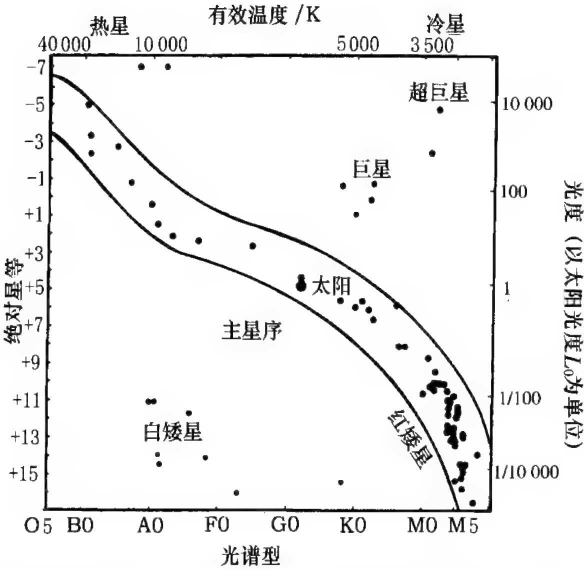
恒星大气的物理特征(温度、压力、密度等)彼此相差很大，按其光谱可分为 O，B，A，R，N 等多种类型.其化学组成与太阳大气差不多，按质量计，氢占 78%，氦占 20%，其他元素占 2%.少数恒星化学组成与太阳不同.

星座(constellation) 天球上的一种分区系统.古人为了认星方便，将星空按亮星排列形状分成若干区域，这种区域即称星座.星座中的众恒星仅表示它们在天球上投影方向互相接近.古希腊人已将北天分成 48 个星座，至今仍被沿用，如大熊座、狮子座等.1603 年后陆续增补了南天星座.1928 年，国际天文学联合会决议全天分成 88 个星座，其中沿黄道带分布的 12 个星座，以春分点开始依次是白羊、金牛、双子、巨蟹、狮子、室女、天秤、天蝎、人马、摩羯、宝瓶和双鱼.

古代将黄道带分成 12 等分，称黄道 12 宫，分别以相应的星座名称命名，如白羊宫、金牛宫等.由于岁差，春分点西退，现在的白羊宫已移到双鱼座.

颜色-光度图(colour-luminosity diagram) 一种天文图.指表示恒星颜色和恒星光度的关系图.该图的横坐标表示恒星颜色，纵坐标表示恒星光度.实际上是赫罗图，常用于星团的研究.

赫罗图(Hertzsprung-Russell diagram) 一种天文图.指恒星光谱型和光度的关系图.丹麦天文学家赫茨普龙(Hertzsprung，E.)于 1911 年和美国天文学家罗素(Russell，H.)于 1913 年各自独立创制，因而简称赫罗图(见图).一般以光谱型(或表面温度)为横坐标，以绝对星等(即光度)为纵坐标，故又称光谱-光度图.大部分恒星分布在自左上角到右下角沿对角线的狭窄带状区(主星序)中，称主序星.其



余则分布在白矮星、巨星和超巨星的几个区域中. 赫罗图将恒星的主要物理参量光谱型（表征温度）和光度结合起来对恒星进行了分类. 用恒星在赫罗图上位置变化可以反映出恒星演化的过程.

主序星（main sequence star） 一类恒星. 处于演化中期的恒星称为主序星. 在赫罗图上这类星位于主星序上，因而得名. 主序星比较稳定，内部产能和向外辐射的能量处于稳定平衡. 银河系绝大多数恒星均为主序星. 其光度与恒星质量成正比，其能源以质子-质子反应或碳氮循环为主，恒星质量和光度越大，恒星在主序星阶段时间越短. 太阳是G2型主序星，它在主序星阶段的寿命约为100亿年.

主序星的质量 M 和光度 L 存在一定的关系——质光关系，即

$$L \propto M^4 \quad (L > L_{太阳});$$

$$L \propto M^{2.8} \quad (L < L_{太阳}).$$

巨星（giant） 一类恒星. 光度级为Ⅲ的恒星称为巨星. 它们在恒星赫罗图上分布在主星序和超巨星之间，光谱型从O型到M型，著名的巨星有五车二（御夫座α）、昴宿六（金牛座η）等. 光度级为Ⅱ的恒星称亮巨星，在赫罗图上分布在超巨星和巨星之间（参见“光度级”）.

超巨星（supergiant） 一类恒星. 光度级为 $\mathrm{I_a}$，$\mathrm{I_b}$ 的恒星称为超巨星. 位于赫罗图的顶部. 绝对目视星等大于−2等. 光谱型从G型到M型，温度范围为3000～30000K. 光度是太阳光度的 10^4～10^6 倍，最大的可达 10^9 倍. 温度高的称为蓝超巨星，肉眼可见的最亮的蓝超巨星为猎户座β（参宿七）、天鹅座α（天津四）；温度低的称为红超巨星，肉眼可见的最亮的红超巨星为天蝎座α（心宿二）和猎户座α（参宿四）. 超巨星体积巨大，最大的其半径为太阳半径的几千倍，如HR237的半径约为太阳半径的1800倍. 超巨星是年轻的、演化迅速的大质量恒星，其寿命较短，具有延伸的大气.

矮星（dwarf star） 亦称主序星. 一类恒星. 指光度级为Ⅴ的恒星. 光度为 10^{-4}～10^4 太阳光度，光谱型为O，B型到M8型，表面温度为2000～35000K，质量为0.1～20个太阳质量，是银河系中最普通的一类星，占总数90%以上. O，B，A型的称为蓝矮星；F，G型的称为黄矮星；K，M型的称为红矮星. 太阳是一个典型的黄矮星. 绝对星等比同光谱型的矮星约暗2星等的恒星，称为亚矮星，它们在赫罗图上位于主星序和白矮星序之间，构成大致与主星序平行的亚矮星序，大都为红色或黄色恒星. 白矮星专指一种半径很小、密度非常大、光度低的恒星（参见“白矮星”）. 黑矮星指小质量恒星演化到晚期不再发出辐射能的恒星，或指质量小于0.08太阳质量、没有核反应能源的恒星. 亚矮星、白矮星、黑矮星均非矮星.

亚矮星（subdwarf） 见“矮星”.

红巨星（red giant） 一类恒星. 光谱为K和M型，光度级为Ⅲ的红色或橙色的恒星称为红巨星. 其温度约为4000K，直径比太阳大几十倍. 现代恒星演化理论认为：恒星演化到主序星后期，中心区的氢燃烧完毕，中心区的重力大于向外的辐射压，核心区开始收缩，其外层变热、变稠，开始氦燃烧，这时外层得到核心释放的能量而剧烈膨胀变成体积很大、密度很稀、表面温度很低、光度很强的红巨星，常伴随物质抛射形成行星状星云. 大多数恒星演化都要经过红巨星这一重要阶段. 太阳已经耗掉了其中心区三分之一左右的氢，当其中心区氢聚变为氦的核反应过程结束时，太阳也将变成红巨星. 肉眼所见的红巨星有大角（牧夫座α，K2Ⅲ）、毕宿五（金牛座α，K5Ⅲ）、北河三（双子座β，K0Ⅲ）.

白矮星（white dwarf） 一类恒星. 光度很低（“矮”）、温度很高（“白”）的电子简并态的恒星称为白矮星. 其磁场强度为10～10^3T，密度高达 10^8～10^{10}kg/m^3. 现已发现1000多颗. 现代恒星演化理论认为：当恒星燃烧尽其热核燃料，剩下的质量小于1.4太阳质量时就会演化为白矮星. 白矮星的有效温度大都介于5500～40000K之间，质量一般在0.2～1个太阳质量，它们在赫罗图上的分布区域位于主星序的左下方. 天狼星伴星是最早发现的白矮星. 由于它们的光度很低，绝对目视星等在8～16等之间，因此，只有离太阳不太远的白矮星才能被看到.

有效温度（effective temperature） 天体物理学术语. 指表征处于热动平衡状态下恒星大气物理性质的一种参量. 即一个与某恒星大小相同的绝对黑体发出的总辐射能与该恒星总辐射能相同时，绝

对黑体所具有的温度. 太阳表面的有效温度为5770K. 恒星的光谱型与其有效温度密切相关,不同光谱型相应的有效温度如下:

O 型: $3\times10^4\sim4\times10^4$K;

B 型: $10^4\sim3\times10^4$K;

A 型: $7.5\times10^3\sim10^4$K;

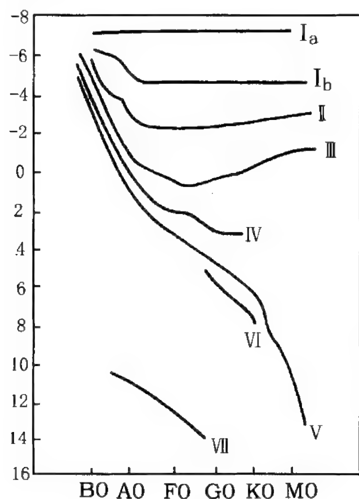
F 型: $6\times10^3\sim7.5\times10^3$K;

G 型: $5\times10^3\sim6\times10^3$K;

K 型: $3.5\times10^3\sim5\times10^3$K;

M(R,N,S)型: $2.5\times10^3\sim3.5\times10^3$K.

光度级(luminosity class) 亦称光度型. 天体物理学术语. 定义为恒星发光强度大小的等级. 20世纪40年代,美国天文学家摩根(Morgen,W. W.)等人把恒星按其光度共分成七个等级,分别记为Ⅰ,Ⅱ,Ⅲ,Ⅳ,Ⅴ,Ⅵ,Ⅶ(如图),相当于超巨星、亮巨星、巨星、亚巨星、主序星、亚矮星和白矮星的光度级. 常缀于恒星光谱型之后,如太阳为G2Ⅴ,表示太阳是光谱为G2型的矮星.



质光关系(mass-luminosity relation) 天体物理学的一种规律. 指恒星质量与光度间的关系. 哈姆(Halm,J. K. E.)最早提出,1919年,由赫茨普龙(Hertzsprung,E.)用观测资料证实. 1924年,英国天文学家爱丁顿(Eddington,A. S.)从理论上导出恒星的光度 L 与其质量 M 存在 $L=kM^{3.5}$ 的简单关系,k 为比例常数. 实测资料得出:对 $L>L_\odot$ 的主序星,

$$\frac{L}{L_\odot}\approx\left(\frac{M}{M_\odot}\right)^4;$$

对 $L<L_\odot$ 的主序星,

$$\frac{L}{L_\odot}\approx\left(\frac{M}{M_\odot}\right)^{2.8},$$

$L_\odot$,$M_\odot$ 分别表示太阳的光度和质量. 除物理性质特殊的巨星、白矮星和一些致密天体外,对占恒星绝大多数的主序星都符合这一质光关系. 利用质光关系可以由光度来估算恒星质量,还可以用作研究恒星内部结构、建立各种理论模型的判据.

自行(proper motion) 天体物理学术语. 指恒星本身的空间运动沿垂直于视线方向的分量在天球上的投影,或恒星空间运动引起的每年在天球上的角度位移. 以每年或每百年在天球上移动的角度表示,单位为(角)秒/年或(角)秒/百年. 恒星自行一般小于0.1″/a. 现在已知自行最大的恒星为蛇夫座中的巴纳德星,其自行为10.31″/a.

沃尔夫-拉叶星(Wolf-Rayet star) 一类特殊的恒星. 指光谱中有很宽发射线的O型和B型星. 1867年,法国天文学家沃尔夫(Wolf,J. E.)和拉叶(Rayet,G. A. P.)在天鹅座发现3颗8等星的光谱中有很多很宽的发射线,因而得名,符号为WR. 迄今已经发现约250颗,其中最亮的是船帆座 γ^2,视星等为1.7等. WR星的光谱是强连续谱叠加中性氦、电离氦及电离的碳、氮、氧的发射线,发射线宽度达几纳米. 由谱线的宽度和轮廓可知,有物质以1000~2000km/s的速度从星体中流出,并在星的周围形成运动着的延伸大气包层. 大多数WR星的光谱可分为两个次型:碳序和氮序. 前者光谱中以电离碳和电离氧谱线为主,后者电离氮谱线很强,两者均有很强的氦谱线,与普通O型星和B型星相比,WR星大气中氢的含量少50~150倍,氮序WR星大气中氮含量高50~100倍,碳序WR星大气中碳含量高400~700倍. WR星是很年轻的恒星,从子星为WR星的双星可以估计WR星的初始质量很大. 由于大质量恒星演化很快,因此WR星已演化到主序星的后期,大部分氢已燃烧转化为氦,故WR星的氦、氢含量比超过正常星的几十倍.

共生星(symbiotic star) 一类特殊的恒星. 其光谱既出现低温星吸收线、又出现高温星发射线的一种恒星称为共生星. 1941年,梅里尔(Merrill,P. W.)首先定名. 其光变具有准周期的类新星爆发特征,并有小幅度的快速非周期光变. 光变幅达3个星等,其光变与谱变有一定的相关. 当光度增强时,晚型吸收谱和高激发发射线减弱或消失,B型气壳谱线增强;当光度减弱时,晚型吸收谱和高激发发射线重新出现或增强. 共生星的空间分布集中在银道面附近,属年老的盘星族. 现在认为共生星是由低温的红巨星和高温B型或O型亚矮星组成的双星,两子星被一共同的气壳包裹着. 目前已知有50颗,典型星是仙女座Z星.

天鹅座P型星(P Cyg star) 一类恒星. 指一种光谱线具有特殊谱线轮廓的恒星. 其发射谱线旁有一条吸收子线,发射线没有位移,吸收子线向蓝端位移. 天鹅座P星是这类恒星的典型代表,因而得名. 1600年发现它突然增亮3个星等,一度被看做是新

星，此后光度有一系列起伏变化，到 1715 年后亮度基本稳定下来.该星具有 BⅠ型超巨星的光谱，但附加有发射线.它的距离为 1.8×10^3pc，绝对星等为－9 等，表面有效温度为 17000K，半径为 60 太阳半径，质量为 40 太阳质量.它不断有物质抛射出，质量损失率为 3×10^{-5}太阳质量/年.抛射出的物质形成向外膨胀的气壳.目前，已发现天鹅座 P 型恒星仅几十颗.它们的光谱型在 O6 和 A4 之间，绝对目视星等从－3 等到－8.5 等，有不规则小幅度光变.它们具有向外不断膨胀的气壳，膨胀速度可达每秒几百千米，每年质量损失为 $10^{-5}\sim10^{-4}$太阳质量.

A 型特殊星(Ap star)　一类恒星.指具有特殊光谱的 A 型星.光谱中一些吸收线的强度呈周期性变化，往往拥有很强而变化的磁场，符号为 Ap.早在 19 世纪末进行恒星光谱分类研究时就发现，一部分早 A 型星的光谱中电离硅、电离锰、电离铬、电离锶、电离铕的谱线特别强，因此得名.在光谱型 B8～F0 的恒星中，大约有十分之一的星为 A 型特殊星.Ap 星磁场强度一般在 0.1T 以上，其中 HD215441 的磁场强度高达 3.4T，并且磁场有周期性变化，其中一部分还有光度和光谱变化.光变周期为 1～25d，变幅一般不超过 0.1 等.典型星是猎犬座 α^2 星(中文名为常陈一)，在 5.469d 内磁场强度从 0.162T 变到－0.142T.亮度变化的 A 型特殊星又称猎犬座 α^2 型变星(常陈型变星).Ap 星的光谱中铬和铕的谱线强度常呈现出有规律的变化.Ap 星属于星族Ⅰ.最亮的 Ap 星是大熊座 ε(中文名玉衡).

B 型发射星(Be star)　一类恒星.指光谱中出现发射线的 B 型星.1866 年，意大利天文学家塞奇(Secchi，A.)首先发现仙后座 γ 的光谱中有氢的发射线，1922 年，国际天文学联合会第一届大会正式命名.最亮的 Be 星是波江座 α(中文名水委一)，视星等为 0.46 等.Be 星的光谱的吸收线和连续谱与 B 型星类似，但吸收线通常很宽，光谱中常出现氢的发射线，发射线强度和形状具有缓慢而不规则的变化.约 70％的 Be 星有 0.1 等的小幅度光变.Be 星都在快速自转，赤道自转速度可达 400～450km/s.现已公认 Be 星有星周包层.

原恒星(protostar)　一类天体.指处于恒星演化阶段极早期的天体.根据现代恒星演化理论，恒星是由弥漫的星际物质通过引力收缩过程形成的.处于引力收缩阶段的浓密星际物质云，称为原恒星.观测发现，某些红外源，如隐藏在分子云内的贝克林-诺伊格鲍尔红外天体(BN 天体)等，极可能是正在形成的恒星，即原恒星.

红外星(infrared star)　一类恒星.指辐射能量集中在红外波段的恒星.表面温度很低，只有几百摄氏度.体积很大，半径为太阳半径的几十甚至几千倍.它们一部分是年幼星或正在从星云中形成的恒星，正处于引力收缩阶段，其温度低，只能发射不可见的红外光；另一部分红外星是年老的将要死亡的恒星，由于核心收缩而释放能量，使外层膨胀成为温度很低的红外星.红外星主要是某些红超巨星、长周期变星、碳星、淹没在浓厚吸光云中的 O，B 型星、B 型发射星、爆发不久后的新星、老新星、金牛座 T 型变星和猎户座 BN 天体等.

星风(stellar wind)　一种天文现象.指从恒星不断向外放射出的物质流.对于太阳，已经直接探测到太阳风的存在.对于 O 型、B 型超巨星，其光谱中普遍存在发射线旁边出现紫移的吸收线，从吸收线的位移可测出视向速度达每秒数千千米，表明这些星的外壳正以高速向外膨胀.对于 M 型巨星和超巨星，其光谱中强吸收线分成两条，一条锐而深的吸收线(形成于光球之外的星周物质中)，相对于另一条宽而浅的吸收线(形成于恒星光球之中)紫移，谱线位移相对速度为 10km/s.此现象表明星周物质在膨胀，一般地，对 O 型、B 型恒星，由于恒星的辐射压和快速自转或受某种扰动而抛射出的带电粒子，主要是氢离子，在磁场作用下，星风以冲击波的形式，以每秒上千千米的高速在星际空间传播.对于 M 型巨星，星风的起源可能是由于某种波(如声波)不断输送能量给恒星外层大气而形成的.星风造成的恒星质量损失率很小，如太阳风每年约损失 2×10^{-14} 太阳质量.

五车二(Capella)　亮星之一.即御夫座 α 星(α Aur).历元 2000.0 年的坐标为：赤经 5h17min，赤纬＋46°00′.视星等为 0.08 等，为一双谱分光双星，轨道周期约 104d，其子星甲星的绝对星等为 0.12 等，是 G 型黄巨星(G5Ⅲ)，表面温度为 5280K，光度为太阳光度的 140 倍，半径为太阳半径的 14 倍，质量为 2.67 太阳质量；子星乙星的绝对星等为 0.37 等，也是 G 型黄巨星(G0Ⅲ)，光度为太阳光度的 60 倍，半径为太阳半径的 8.9 倍，质量为 2.55 太阳质量.已发现它有 X 射线辐射.

牛郎星(Altair)　亮星之一.即天鹰座 α 星(α Aql).历元 2000.0 年的坐标为：赤经 19h51min，赤纬＋8°52′.视星等为 0.77 等，绝对星等为 2.19 等，距离地球 16 光年.为一 A 型白色主序星(A7Ⅴ)，表面温度为 8010K，光度为太阳光度的 10.5 倍，半径为太阳半径的 1.68 倍，质量为 1.6 太阳质量.可能是物理双星，轨道周期初步值约为 12.5a，伴星的质量约为 0.05 太阳质量.观测发现牛郎星在快速自转，每天至少转 2 周.

大角星(Arcturus)　亮星之一.即牧夫座 α 星(α Boo).历元 2000.0 年的坐标为：赤经 14h15min，

赤纬$+19°11'$. 全天第四亮星，视星等为-0.04等，绝对星等为-0.23等，距离地球35光年. 为一K型红巨星(K2Ⅲ)，表面温度为4400K，光度为太阳光度的190倍，半径为太阳半径的23倍，质量为0.6太阳质量(目前仍未定准). 大角星的一个有趣特性是：对太阳而言空间速度很大，视向速度为+90km/s，在天球切面观测到的自行相当于120km/s. 其光谱中CN带较正常K型巨星为弱，氧含量出奇地高. 它属于星族Ⅱ.

老人星(Canopus)　亮星之一. 即船底座α星(α Car). 历元2000.0年的坐标为：赤经6h24min，赤纬$-52°42'$. 全天第二亮星，视星等为-0.73等，绝对星等约-4.7等，为一白色F0型超巨星(F0 I_b)，表面温度为7460K，光度为太阳光度的6000倍，半径为太阳半径的46倍，质量为太阳质量的12倍，距离地球200光年，是一颗磁变星，磁场强度变化于$-10^{-2}\sim6\times10^{-2}$T，变化周期为22.35d，有X射线辐射.

天狼星(Sirius)　亮星之一. 即大犬座α星(α CMa). Sirius为希腊语，意为"白狗"、"牝犬". 历元2000.0年的坐标为：赤经6h45min，赤纬$-16°43'$. 为一目视双星，轨道周期为50a，主星天狼A为全天目视第一亮星，视星等为-1.46等，绝对星等约1.43等，为一白色A型主序星(A1Ⅴ)，表面温度为9970K，光度为太阳光度的25倍，半径为太阳半径的168倍，质量为太阳质量的2倍，是一金属线星，距离地球8.65光年. 其伴星(天狼B)为一白矮星，常称天狼伴星(参见"天狼伴星"). 已发现主星和伴星均有X射线辐射. 公元前30世纪，古埃及人已知每当天狼星同太阳东升(偕日升)，尼罗河即发生泛滥，据此把天狼星偕日升的周期定为一年，那一天定为岁首，这是人类最早的"太阳历".

天狼伴星(companion of Sirius)　天狼星的伴星. 1844年，德国天文学家贝塞尔(Bessel，F. W.)根据天狼星移动轨迹中出现的波浪形，推断它必定有一颗没有看见的伴星存在. 1862年，美国光学家克拉克(Clarck，A. G.)用他自己制造的当时最大口径47cm的折射望远镜看到了它，视星等为8.68等，光度为太阳光度的0.03倍，半径为太阳半径的0.0073倍，质量和太阳质量相当(1.05太阳质量)，密度达3.8×10^9kg/m^3，密度是水的密度的3.8×10^6倍，表面温度为26000K. 它是最早发现的白矮星，与天狼星组成双星，双星绕转周期为50a.

参宿四(Betelgeuse)　亮星之一. 即猎户座α星(α Ori). 历元2000.0年的坐标为：赤经5h55min，赤纬$+7°24'$. 视星等为0.06～0.75等，全天最亮的红超巨星(MⅠ)绝对星等约-6等，是一颗半规则变星. 表面温度约3500K，光度约为太阳光度的10^5倍，半径为太阳半径的900倍，质量为太阳质量的15倍，距离地球约600光年. 1966年，发现它是一射电星，还发现它周围有极厚的气壳，延伸到本星半径约600倍处.

南门二(Rigil Kent)　亮星之一. 即半人马座α星(α Cen). 历元2000.0年的坐标为：赤经14h39min，赤纬$-60°50'$. 全天第三亮星，视星等为-0.27等，距离地球4.35光年. 是三合星，甲、乙两星组成目视双星，轨道周期约80a. 甲星为黄色G型主序星(G2Ⅴ)，视星等为0.01等，绝对星等为4.38等，表面温度为5800K，光度为太阳光度的1.51倍，半径为太阳半径的1.2倍，质量为1.10太阳质量. 乙星为红色K型主序星(K1Ⅴ)，视星等为1.34等，绝对星等为5.72等，表面温度为5300K，光度为太阳光度的0.47倍，半径为太阳半径的0.84倍，质量为0.91太阳质量. 现已分别观测到两星的X射线. 丙星又称比邻星，是距地球最近的恒星，距离地球4.22光年(参见"比邻星").

比邻星(Proxima)　近星之一. 半人马座α三合星的子星丙星. 除太阳外是离地球最近的恒星，距离地球4.22光年. 视星等为11.05等，为一亮度有变化的耀星，绝对星等为15.49等，是一红色矮星(dM5e). 表面温度为2700K，光度为太阳光度的0.0017倍，半径为太阳半径的0.19倍，质量约为0.11个太阳质量(参见"南门二").

织女星(Vega)　亮星之一. 即天琴座α星(α Lyr). 历元2000.0年的坐标为：赤经18h37min，赤纬$+38°47'$. 视星等为0.03等，绝对星等为0.50等，距离地球26.3光年，为一A0型主序星(A0Ⅴ)，表面温度为9660K，光度为太阳光度的60倍，半径为太阳半径的2.76倍，质量为太阳质量的2.4倍，它是光度测量、光谱分类的标准星，是一颗变星，光变幅达0.08等. 近年来，空间探测还发现它有X射线辐射.

心宿二(Antares)　亮星之一. 即天蝎座α星(α Sco). Antares源自希腊语，意为"火星的劲敌"，指它的颜色和亮度与火星相近似. 历元2000.0年的坐标为：赤经16h29min，赤纬$-26°26'$. 视星等为0.86～1.02等，绝对星等约-5.3等，是有名的红超巨星(M1Ⅰ). 表面温度为3600K，光度为太阳光度的5×10^4倍，半径为太阳半径的600倍，质量为太阳质量的25倍，距离地球410光年. 是一目视双星，其伴星为B5型蓝矮星，视星等约5等，距离主星约300天文单位.

巴纳德星(Barnard star)　一颗恒星. 指蛇夫座中一颗具有大自行的恒星. 1916年，美国天文学家巴纳德(Barnard，E.)发现它的自行达10.31″/a，是已知自行最大的星，因而得名. 它是一颗M5Ⅴ型红

矮星，距离地球 5.9 光年，是除太阳和南门二以外距地球最近的恒星. 视星等为 9.54 等，其视向速度为 -108km/s，切向速度为 90km/s，空间速度为 140km/s. 有的观察分析认为它有甲乙两个伴星，甲星轨道周期为 11.7a，相对巴纳德星的轨道半径约 2.7 天文单位，质量为 0.8 个木星质量. 乙星相应的数据分别为 20a、3.8 天文单位、0.4 个木星质量. "高能天文台 2 号"探测器已测得它有 X 射线辐射.

北极星(Polaris)　亮星之一. 指距北天极(北天极是地球自转轴交天球上的点)最近的亮星，现为小熊座 α 星(α UMi，中文名为勾陈一或北辰). 它距北天极约 1°. 北斗七星斗魁末端的两颗亮星指向北极星，北极星实际为一三合星，主星甲是离地球最近的造父变星，亮度变化于 1.95～2.04 等，光变周期为 3.97d，是一黄白色的 F 型超巨星(F8 I_b)，光度为太阳光度的 3000 倍. 它又是轨道周期为 30a 的分光双星，伴星乙视星等为+8.6 等，距离主星甲 18″，距离地球 680 光年. 由于地轴运动，北天极以约 26000 年的周期围绕黄极运动，一些离北天极较近的恒星依次取得北极星的称号. 公元前 2750 年，天龙座 α 星(中文名为右枢)曾是北极星. 目前，北天极以 15″/a 的速度向小熊座 α 靠拢，公元 2100 年相距 28′. 2000 年以后，仙王座 γ 星(γ Cep)将成为北极星. 公元 7000 年、10000 年，14000 年，北极星依次为仙王座 α、天鹅 α、天琴座 α.

双星(binary star)　一种天体. 指靠得很近的两颗恒星. 主要指空间距离很近、成为一个力学系统的物理双星. 有时把两颗单星在空间相距很远，彼此独立，仅是在天球上的投影位置彼此很接近的现象称为光学双星或视双星. 物理双星在相互引力作用下绕公共重心旋转. 可分为目视双星、分光双星、交食双星、密近双星等. 目视双星是通过望远镜用目视或照相方法可以分辨出两颗子星的物理双星，两子星彼此绕转周期较长，一般在百年到千年甚至万年，短的约 1 个月，大熊座 ζ(中文名开阳)是第一颗被发现的目视双星. 分光双星是用分光方法测定其谱线的多普勒位移来判知的双星，两子星距离很近，用大的天文望远镜都不能分辨为双星. 交食双星也称食双星、光度双星、食变星，为两子星相距较近、互相绕转的轨道面与视线方向倾角很小的物理双星，由于绕转时引起掩食，使亮度作周期性变化，故名. 英仙座 β、天琴座 β 是典型的食双星. 密近双星是两子星相距甚近，甚至几乎接触、彼此间常有物质交流的双星，密近双星的两子星相互绕转周期短，多数约为半日.

物理双星(physical double)　见"双星".

视双星(optical double)　亦称光学双星. 一种天文现象. 指两颗彼此无物理联系、互不相关的、远看互相接近但实际相距很远的恒星. 视双星彼此没有公转运动和公共自行，平均视向速度也不相等. 区分物理双星和视双星，需要大量的观测和分析. 物理双星彼此互相绕转，两星运动轨迹为一条曲线或一条椭圆曲线. 对周期短的物理双星，经过一定时期的观测即可决定；对周期长的双星，需要研究两子星的自行才能决定. 如果两子星的自行相同，则为物理双星；如果两子星的自行不同，则为视双星.

光学双星(optical double)　即"视双星".

目视双星(visual binary)　一种物理双星. 指通过望远镜能直接用肉眼或用照相方法分辨出是两颗子星的物理双星. 两星彼此绕转，周期较长，绝大多数的周期超过几年，短的 1 个月，长的达几百年以上. 第一对目视双星是意大利天文学家利齐奥里(Riccioli，J. B.)于 1650 年发现的，它就是大熊座 ζ(开阳). 目视双星的主星与伴星的角距离和位置角(主星和伴星的连线与南北方向的夹角)，用附在望远镜上的测微器测定，也可用照相方法和干涉方法测定. 用测定的数据可求得子星绕转的视轨道和真轨道参数，进而可求得子星的真轨道. 由观测得到轨道周期 P 和轨道半径 a，则由开普勒第三定律可求出双星的质量和 M_1+M_2，即

$$\frac{a^3}{P^2}=\frac{G}{4\pi^2}(M_1+M_2).$$

若测定出每颗子星的绝对轨道的半长径 a_1 和 $a_2=a-a_1$，进一步可求出子星的质量比，即

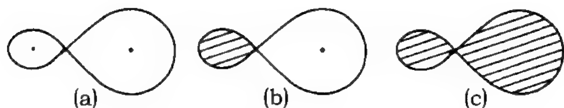
$$\frac{M_1}{M_2}=\frac{a_2}{a_1},$$

从而计算出每个子星的质量. 观测目视双星是测定恒星质量的可靠和直接的方法. 现已发现的目视双星约 8 万对.

分光双星(spectroscopic binary)　物理双星之一. 利用分光方法从子星的视向速度变化来判定的物理双星称为分光双星. 由于两颗子星靠得很近，用大的天文望远镜也难以分辨，常用分光的方法测定谱线的多普勒位移来求得两子星的真正速度. 星对愈紧密，两星的轨道周期愈短，绕转速度愈快，多普勒位移就愈大. 视向速度变化范围为每秒几千米到每秒数百千米. 两子星的谱线都已测得的称为双谱分光双星，只测得一个子星谱线的称为单谱分光双星. 如果两子星间的距离远，因而轨道速度小，或者双星轨道面的法线与视线的交角很小，观测不到谱线的多普勒位移，这时如果两子星的光谱型相差很大，可以根据光谱是由两个光谱组成来判断是双星，这种双星称为光谱双星. 已发现的分光双星约有 5 千对. 轨道周期最短的分光双星是天箭座 WZ，为 82min. 最长的为蛇夫座 70，周期长达 88a. 亮星五车二、角宿一、大陵五等都是分光双星.

交食双星(eclipsing binary)　物理双星之一. 其轨道面的法线与视线的交角接近 90°的物理双星称为交食双星. 由于子星相互位置变化,发生互相掩食,使亮度发生周期性的变化. 根据光变曲线的形状可分为三种类型:大陵五型食双星、渐台二型食双星和大熊座 W 型食双星. 交食双星又称食变星,常编入变星表中,已知有 4 千多对. 据估算,在银河系中交食双星可达几亿对. 某些交食双星又是双谱分光双星,它们能提供可靠的恒星基本参数.

密近双星(close binary)　物理双星之一. 指两子星相距很近,甚至几乎接触,彼此间常有物质交流的双星称为密近双星. 密近双星的两子星互相绕转很快,周期长的只有几天,多数约为半天. 已知周期最短的是天箭座 WZ,轨道周期只有 88min. 1955 年,科帕尔(Kopal,Z.)提出按引力等位面的双星分类法,将双星分为三类:不相接双星,两子星都未充满等位面(两子星都比各自的洛希瓣小);半相接双星,只有一颗子星充满等位面(一个子星完全充满或几乎充满洛希瓣);相接双星,两子星都充满等位面(两子星都完全充满或几乎充满洛希瓣). 密近双星中两子星的质量交流会导致轨道运动周期发生变化、视向速度曲线和光变曲线出现畸变等,观测密近双星可研究恒星内部结构状况、星周物质及质量交流、星风、吸积等.



不相接(a)、半相接(b)和相接(c)双星

聚星(multiple star)　恒星系统之一. 由三颗至 10 多颗恒星在相互引力作用下组成的恒星系统称为聚星. 由三颗恒星组成的恒星系统又称三合星,四颗星在一起的称四合星,依次类推. 大熊座 ζ(中文名开阳)就是个六合星. 大熊座 ζ 和大熊座 81 星(中文名为辅)组成一双星系统,彼此相距 19000 天文单位,大熊座 ζ 本身是由相距 400 天文单位的两对双星组成,辅也是分光双星. 聚星可分为两类:一类为普通聚星,如大熊座 ζ,成员星由双星组成,成员星之间的距离远大于双星间的距离,成员星的运动类似双星的周期运动. 另一类是四边形聚星,是由四颗星组成的恒星系统,成员星之间的距离相差不多,成员星的运动不是周期性的,这种四边形聚星系统是不稳定的,主要存在于星协中,为星协核心的一部分,如猎户座星协的核心就是猎户座四边形聚星.

变星(variable star)　一类恒星. 亮度(或电磁辐射)经常变化的恒星称为变星. 亮度变化的范围可从千分之几等到 20 等,甚至更大. 其中有的是双星,由于双星轨道平面与视线方向重合,双星运动引起交食而产生亮度变化的称为食变星或几何变星;由于恒星本身或它周围大气的实际物理变化而引起亮度(或电磁辐射)变化的称为物理变星. 物理变星可分为脉动变星和爆发变星两大类,根据光变的形态和物理原因又细分为若干次型,如造父变星、金牛 T 型变星、天琴座 RR 型变星、蒭藁型变星、新星、超新星等. 中国河南殷墟出土的商代《甲骨卜辞》中有“新大星并火”和“有毁新星”的记载,可认为是最早的爆发变星记录. 1596 年,法布里修斯(Fabricius,J.)最早发现鲸鱼座 o(蒭藁增二)的亮度有变化,后来确定它的亮度变化有约 11 个月的周期. 1667 年,蒙塔纳里(Montanari,G.)发现英仙座 β(大陵五)、1784 年,古德里克(Goodricke,J.)发现仙王座 δ(造父一)是变星. 1900 年发现的变星总数达 700 颗. 由于观测技术的发展,20 世纪发现变星的数目剧增. 1985 年后,陆续出版的《变星总表》(第四版)收录了到 1982 年为止发现和命名的变星 28450 颗. 变星发现后,如已有专名,仍保留专名,否则按国际通用命名法给以命名,即每个星座内的变星按变星发现的次序,依次用拉丁字母(J 不用)R,S,…,Z,RR,RS,…,RZ,SS,ST,…,ZZ,AA,AB,…,QZ(共 334 个),加上星座名称为为变星的符号. 如天琴座 RR 变星,该星座内第 335 个以后的变星开始用 V335,V336 等加星座名称来命名.

几何变星(geometric variable star)　亦称光学变星. 一类变星. 指由于几何原因(双星交食、自转)造成亮度变化的变星. 双星的两个子星在互相绕转时,当双星轨道面和视线方向近于平行时,一子星会掩食另一子星,使亮度发生变化,故又称食变星,或食双星. 有的椭球双星公转时两子星的最长径总在一直线上,交食时,两星最长轴差不多和视线平行,星的视面最小,使亮度变小;交食之间最长径大致垂直于视线,星的视面最大,从而亮度变大. 食变星可分为大陵五型(因典型星为英仙座 β 星(中文名为大陵五)而得名);渐台二型(因典型星为天琴座 β 星(中文名为渐台二)而得名);大熊座 W 型. 后者是距离极近的相接双星,由于两子星间的质量交流而引起亮度变化. 从几何变星的光变曲线和视向速度变化曲线能够计算出两子星的质量、大小和密度,还能了解其子星间的质量交流情况,有助于研究恒星内部结构和恒星的演化.

物理变星(physical variable star)　一类变星. 因恒星自身内部原因(如脉动、爆发)引起亮度变化的恒星称为物理变星. 可分为体积膨胀、收缩的脉动变星和亮度突然增大的爆发变星. 每类又按光度和光谱变化特征分为许多次型. 近年来,把光学波段的物理条件有变化的(如光谱变星、磁变星)和其他电

磁波段电磁辐射有变化的恒星(如红外变星、X 射线新星等)都称为物理变星. 各类物理变星在赫罗图上分布在主星序和红巨星之间的变星区,研究物理变星可为主序星前期和恒星晚期的演化提供线索. 此外,物理变星在空间分布上分属于不同星族,研究物理变星的星族特性也为研究银河系的结构和银河系动力学提供重要线索.

脉动变星(pulsating variable) 亦称大序变星. 物理变星中的一种. 其亮度变化可用其体积的时而膨胀、时而收缩的脉动理论来解释,故名. 星体交替地膨胀和收缩,引起星的半径、光度、温度和视向速度等变化. 大多数变星都属于此类,在银河系内已发现 1 万多颗. 这类变星的特点是存在周光关系(光度和变光周期的对数成正比)和周谱关系(变光周期越长,光谱型越晚). 按变光周期长短、规律性和在银河系内的分布,可分为长周期造父变星、短周期造父变星、矮造父变星、长周期变星、半规则变星、不规则变星六类. 还可细分为经典造父变星、室女座 W 型变星、天琴座 RR 型变星、仙王座 β 型变星、盾牌座 δ 型变星、蒭藁型变星、半规则变星、金牛座 RV 型变星、不规则变星等类型. 长周期造父变星又称造父变星,因其典型星为仙王座 δ 星(中文名为造父一)而得名. 又可分为经典造父变星和室女座 W 型变星. 短周期造父变星,因典型星为天琴座 RR 星,又称天琴座 RR 型变星,因在星团中发现,旧称星团造父变星或星团变星. 其另一次型为仙王座 β 型变星,又称大犬座 β 型变星,光变幅约为 0.1 等,周期在 0.1～0.3d 之间,存在周光关系. 矮造父变星,包括盾牌座 δ 型变星和船帆座 AI 型变星,光变周期为 0.05～0.3d,都是矮星. 长周期变星,因其典型星为蒭藁增二(鲸鱼座 o),又称蒭藁型变星. 半规则变星的亮度变化有一定周期性,但光变曲线形状和光变周期都有很大的不规则性,周期从 30d 到几年,是 F 型、M 型的巨星或超巨星,光变幅多为 1～2 等. 金牛座 RV 型变星,光变曲线呈双波形状,但亮度极小的深度不固定,光变幅有时达 3 等,周期为 30～150d,光谱型大多为 G 型和 K 型,光谱中常出现氢发射线和氧化钛的吸收带. 不规则变星的变光完全没有规则,大多是 M 型、N 型的巨星或超巨星. 脉动变星在赫罗图上大都分布在位于主星序上方的不稳定区域内. 据认为,它们是恒星演化到短暂的不稳定阶段的产物.

造父变星(cepheid variable) 脉动变星的一种. 是长周期造父变星的简称,其典型星为仙王座 δ 星(中文名为造父一). 1784 年,英国天文学家古德里克(Goodricke,J.)首先发现仙王座 δ 星的亮度变化,这一类变星因此得名. 它们的光变周期为 1～50*d*,光谱型在 F4 型与 G6 型之间改变,目视星等变化于 0.5～1.5 等之间. 存在着周光关系,即变光周期越长,光度越大,光度与周期的对数成正比. 测得造父变星的变光周期,利用周光关系可以得到变星的光度(绝对星等 M). 进一步,由目视星等 m 可以计算出该星与地球的距离 r(由 $M=m+5-5\lg r$). 因此,造父变星又被称为量天尺. 这类变星可分为两个次型:经典造父变星和室女座 W 型变星,前者为大部分,属于星族 I;后者属于星族 II. 目前已知有 500 多颗.

天琴座RR 型变星(RR Lyr variable star) 一种脉动变星. 指短周期的脉动变星. 因典型星天琴座 RR 星而得名,又称短周期造父变星. 最初是在球状星团中发现,因此又称星团变星. 光变周期为 0.05～1.5d,光变幅为 0.5～1.5 等,它们的光谱型大多数为 A 型,一小部分为 F 型. 光变曲线有三种形状,据此又分为三种次型:天琴 RRa,RRb,RRc. RRa 和 RRb 型光变周期较长,平均为 0.5～0.7d,光变曲线不对称. RRa 型亮度上升很快,下降较慢;RRb 型变幅较小,亮度极大,较平坦;RRc 型的光变周期较短,平均为 0.3d,光变曲线几乎对称,光变幅更小. 天琴座 RR 型变星的光度弥散很小,平均绝对星等 M 为 0.6 等,与光变周期无关. 当测出其视星等 m 后,据其绝对星等($M=0.6$ 等)可求出其距离 r(利用公式:$M=m+5-5\lg r$. 考虑星际吸光 A,上式为

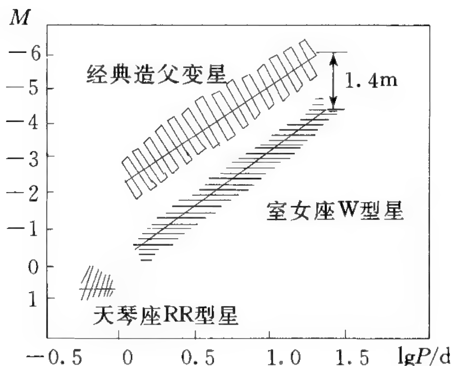
$$M=m+5-A-5\lg r).$$

此外,其质量和半径弥散也很小,平均分别为 0.5 太阳质量、5 倍太阳半径. 已知这类变星的数目约占脉动变星总数的四分之一,但因其光度小,仅在 250kpc 以内能看到它们.

周光关系(period-luminosity relation) 天文学的一种规律. 指造父变星的光变周期和光度之间的一种关系. 1908～1912 年,美国天文学家勒维特(Leavitt,H. S.)研究小麦哲伦星云中的 25 个造父变星时,发现它们的视星等几乎线性地随周期的对数的增加而减小. 它们均属于同一星系,可以认为它们与地球的距离相同,因此周期和视星等的关系反映了周期和绝对星等(光度)的关系,简称周光关系. 对银河系的经典造父变星,周光关系可表示为:

$$M_V=-3.425\lg P+2.52(B-V)-2.459,$$
$$M_B=-3.425\lg P+3.52(B-V)-2.459,$$

式中 M_V 和 M_B 分别表示一个光变周期中绝对目视星等和绝对照相星等的平均值,V 和 B 分别表示经过星际消光改正后的目视星等和照相星等的平均值,P 是以天为单位的光变周期. 后来发现属于不同星族的造父变星,其周光关系不同,属星族 II 的造父变星室女座 W 型变星的光度比经典造父变星小 1.4 等. 对于河外星系的造父变星,其周光关系和银



河系内的造父变星也有所不同.由视星等和周期的关系转化为绝对星等和周期的关系,称为确定周光关系零点问题.假若能测定一颗造父变星的三角视差(即测定距离),并且能估计星际吸光,就不难确定这零点.但造父变星距离太阳相当远,没有一颗能用三角视差法测定其距离,故近半个世纪以来,从事确定周光关系的零点的研究受到挫折.近些年来,在疏散星团和星协中发现了一些造父变星,由于疏散星团和星协的距离能可靠地测定,才使周光关系的零点校正问题得到改进.利用周光关系,由测得的光变周期 P,可求出造父变星的绝对星等 M,再根据它们的视星等 m,利用 $M=m+5-5\lg r$,便可求出含有造父变星的星团或河外星系与地球的距离 r.经典造父变星、室女座 W 型星和天琴座 RR 型星的周光关系可从观测资料得出.

长周期变星(long period variable star)　变星的一种.变光周期为 70～700d 的脉动变星称为长周期变星.是已知变星中数目最多的一种.按光变幅大小、光变周期长短、空间分布和光谱特征分为长周期变星和蒭藁型变星两类.前者光变幅不超过 2.5 等,后者因典型星为鲸鱼座 o 星(中文名为蒭藁增二)而得名.其光变幅超过 2.5 等,个别的甚至可达 10 等.当前,天文学界一般把长周期变星理解为蒭藁型变星.大部分长周期变星的光变周期和光变幅有不规则变化,偏离平均值达 15%,光变曲线各种各样,有的很对称,有的上升比下降快,有的出现双极大.它们是晚型巨星和超巨星,光谱大都属 M 型,少数属于 S,N 和 R 型,光谱中有发射线.长周期变星目视星等变幅较大,部分原因是由于当星的亮度极小时,有效温度最低,大部分辐射从可见光波段转移到红外波段所致.

爆发变星(eruptive variable)　一种变星.由一次星体本身爆发或多次周期性爆发引起光变的变星称为爆发变星.其光变与物质抛射有关.按其爆发规模和性质可分为灾变变星(又称超新星)、激变变星(包括新星、再发新星、类新星、矮新星,矮新星又分为双子座 U 型变星和鹿豹座 Z 型变星)、耀星(又称鲸鱼座 UV 型变星)以及猎户座变星(包括金牛座 T 型变星及御夫座 RW 型变星).狭义的爆发变星通常指激变变星,即指新星、再发新星、类新星、矮新星.已经发现有些爆发变星是密近双星,由于两子星间的物质交流引起爆发.

蒭藁型变星(Mira-Ceti variable)　一种变星.指光变幅大于 2.5 等的长周期变星,也称长周期变星.因典型星鲸鱼座 o 星(中文名为蒭藁增二)而得名.1596 年,法布里修斯(Faburicius,J.)首次注意到它的亮度变化.除新星、超新星之外,它是最早发现的变星.它的光变周期在 320～370d 内变化,亮度极大时可达 1.7 等,亮度极小时在 10 等左右,光变幅达 8 个星等,光谱型在 M5～M9 之间变化.一些蒭藁型变星的资料如下表所示.

星　　名	平均周期/d	目视星等变幅	光　　谱
狮子座 R	312.57	5.1	M7e～M9e
巨蟹座 R	362.06	5.6	M6e～M8e
长蛇座 R	386.2	6.0	M6e～M9e
天鹰座 R	300.3	6.3	M5e～M8e
小狮座 R	372.34	6.9	M7e～M8e
天鹅座 R	426.32	7.7	S4e～S7e
鲸鱼座 o	331.62	8.1	M5e～M9e
天鹅座 χ	406.95	10.9	S7e～S9e

超新星(supernova)　一类爆发变星.即爆发时星等增大 20 等的一种爆发变星.最大光度为新星的 10^4 倍以上,可分为Ⅰ型和Ⅱ型两类.Ⅰ型超新星在光度极大时绝对星等达到 −18 等到 −19 等,20～30 天内下降 2～3 等,此后亮度依指数函数减小,具有强而宽的光谱带.在其光度极大时,光谱中无氢的巴耳末线(Ⅱ型有氢线).Ⅰ型又根据光度极大时光谱有无 SiⅡ的 6150Å 线分为 I_a(有此线),I_b(无此线,但有氦 5876Å 线),I_c(无此线,亦无氦 5876Å 线)三个次型.Ⅱ型超新星最大光度为 −17 等到 −18 等,亮度随时间变化不规则,光谱为连续光谱,紫外光很强.Ⅱ型又根据光变曲线分为Ⅱ-L 型(线性下降)和Ⅱ-P 型(光极大后呈现一平台,然后再下降).超新星是恒星演化晚期阶段所发生的大规模爆发现象,结果使恒星瓦解成星云或抛掉大部分质量后坍缩成为白矮星(I_a 型超新星)、中子星(Ⅱ型超新星)或黑洞.超新星遗迹常是射电源、X 射线源和 γ 射线源.1998 年止已发现河外星系超新星 1445 颗,银河系超新星 8 颗.

新星(nova)　一种爆发变星.爆发时星等增大 7～16 等的变星称为新星.爆发变星爆发时,光度在几小时到两三天内急剧增大上万倍乃至百万倍,以

后缓慢衰减直至恢复原来的亮度.新星并非"新"诞生的星,在爆发前其亮度微弱,直到亮度突然增大,才被"新"发现,实际是恒星演化到晚期阶段所发生的爆发现象.早在公元前14世纪,中国商代的甲骨文上已有新星的记录,西方国家最早的新星记录是公元前131年古希腊天文学家喜帕恰斯(Hipparchus,(R.))的记载.在中国《汉书·天文志》也有详细的记录:"元光元年客星见于房".迄今为止在银河系中共记录了200次新星.

新星的光变特征是:亮度最初增大十分迅速,光变幅度很大,通常达12～13等.在亮度达极大前有短暂停滞,极大过后亮度迅速下降,不久亮度略有起伏,以后亮度缓慢下降.一次新星爆发大约辐射出10^{38}～10^{39}J的能量,其光谱随亮度变化也有变化.亮度极大前相当于B型和A型光谱,吸收线向蓝端位移,相应速度为1000km/s,亮度极大时与A型或F型超巨星光谱类似.亮度减小时出现第二套吸收线,蓝移所相应的速度达1500km/s,有发射线并有氮、氧、铁元素的离子产生的禁线,新星爆发时有气壳向外抛出,抛射速度达1000km/s以上,抛射质量大约为10^{-5}～10^{-3}太阳质量.现在普遍认为,新星是由一颗体积很小、密度很大的白矮星和一颗较冷的主序星或巨星组成的密近双星.冷星的外层气体在白矮星的潮汐力作用下飞向白矮星表面,并使白矮星表面增温,当聚积足够多的氢时,足以产生氢核聚变,使白矮星外层爆发形成新星.

再发新星(recurrent nova) 一类新星.观测到不止一次爆发的新星称为再发新星.与经典新星并无根本区别,可能所有新星均为再发新星.再发新星是爆发周期最短的新星,已观测到十几颗,其中罗盘座T和人马座V2905已各有5次爆发记录.再发新星爆发时,光变幅约在7～8等,比新星的爆发规模小.每次爆发总共释放10^{36}～10^{37}J的能量,抛向星际空间的物质约为10^{-6}太阳质量.

矮新星(dwarf nova) 一类爆发变星.爆发规模较小、爆发周期较短的一种爆发变星称为矮新星.发亮的突然和迅速,与新星和再发新星相似.不过光变幅较小,在2～6等之间,在一天内亮度迅速增亮达到极大,极大后约2～10d亮度很快下降恢复爆发前的亮度.爆发频次较高,爆发平均周期约10～200d不等.按光变曲线的形状,可分为双子座U型变星(或天鹅座SS型变星)和鹿豹座Z型变星两个次型.大多数矮新星属于前者,已发现约300颗.鹿豹座Z型变星比双子座U型变星的光变幅平均小2～3等,周期更短(10～20d),已发现40多颗.

类新星(nova-like variable) 一类爆发变星.指类似新星的爆发变星.光变幅约3个星等,比新星和再发新星小得多,每隔数年亮一次,亮度下降时光谱中出现发射线和禁线,有气壳抛出.一部分类新星是爆发后的老新星,另一部分是共生星(即光谱中既有冷星的吸收线,又有热星的连续发射,还有气壳的高激发发射线的恒星).

金牛座T型变星(T Tau star) 一种不规则的爆发变星.1945年,乔伊(Joy,A.H.)观测到金牛座暗星云中11颗光谱中有H_α发射线的变星,典型的为金牛座T,因此得名.它们具有非周期的不规则光变,或快速光变叠加在长期的缓慢光变上,光变幅一般为1～3等,最大达5等.其光谱为F,G和K型,但在吸收线光谱上叠加氢、电离钙、电离钛、铁等元素的发射线,在光谱的紫区有连续发射.光谱中还出现氧、硫、铁和钛产生的禁线,这些谱线只能在物质密度十分低的区域中,如星云中产生.观测表明,金牛座T型变星伴有弥漫星云,并且成团出现,成为T星协的主要成员.其光谱中吸收线大多比较宽,说明有较大的自转速度,约50km/s.许多金牛座T型变星谱线有蓝移,表明星在向外抛射物质,抛射速度为300km/s,质量损失率估计为10^{-9}～10^{-7}太阳质量/年.其大气内锂丰度比太阳的要大2个量级.锂的过分丰富、总和星云在一起,以及组成T星协等事实,表明它们是非常年轻的恒星,一般认为它们是处在引力收缩阶段的主星序前的恒星.

耀星(flare star) 一类特殊的爆发变星.其亮度在几分钟甚至几秒钟内突然增加零点几等到几等.个别可达10等,经过几分钟到几十分钟恢复到正常状态.早在1924年,赫茨普龙(Hertzsprung,E.)发现船底座DH的亮度突然增大2等,当时未引起注意.1948年,雷登(Luyten,W.J.)发现鲸鱼座UV星在3min内亮度增加12倍.1959年,国际天文学联合会正式把这种亮度增加率达到0.3星等/分钟以上的亮度突变(称为耀亮)的星称为耀星.由于典型星是鲸鱼座UV星,故又称鲸鱼座UV型变星.在太阳附近已发现了约100颗,其中一半是双星的子星,它们都是光度非常低的红矮星,光谱型绝大多数属M型.在射电和X射线波段也观测到耀亮.耀亮的频数与星的光度有关,光度越小的星耀亮越频繁,耀亮发生在占星面1%的区域里,这和日面耀斑现象相似.但两者差别很大,太阳耀斑释放的能量不超过太阳正常辐射能的1%,而耀亮发生时,释放的能量可达耀星正常状态辐射能量的10倍.耀亮是年轻恒星特有的现象.近年来,对个别早型星和大熊座W食双星等也发现了耀亮活动,它们常被称为非经典耀星.

星际物质(interstellar matter) 一种天体.指银河系中存在于星际空间的物质,包括星际气体、星际尘埃和宇宙线.星际物质总质量约占银河系总质量的10%,密度不均匀,介于10^{-31}～10^{-26}g/m^3.星

际气体和尘埃聚集密度较大的区域就成为星云.星际物质集中在银道面,尤其集中在银河系旋臂中.星际气体包括气态原子、分子、电子、离子.其中氢占90%,氦约占10%,其他元素占不到1%.根据氢的存在形式,星际气体的分布可分为电离氢区(HⅡ区)和中性氢区(HⅠ区).星际尘埃是直径0.1～100μm的固态微粒,可能是水冰、甲烷冰、氨冰等微粒,以及二氧化硅、三氧化二铁等矿物粒子和石墨晶粒.星际尘埃散射星光使星光减弱,称星际消光.星光颜色随之变红,称星际红化.

星云(nebula) 一种天体.指银河系内由气体和尘埃组成的星际云.18世纪,英国天文学家赫歇耳(Herschel,J. F. W.)等人把天空中发现的云雾状天体统称星云.后来发现许多星云是在银河系之外,曾将星云分为河外星云和河内星云.现在把河外星云称为河外星系或星系,它是和星云性质截然不同的天体;河内星云又称为银河星云或星云.按发光性质,星云可分为发射星云、反射星云和暗星云;按形状可分为无定形弥漫星云和圆形(或圆环形)的行星状星云、超新星爆发后的残存物质云(超新星余迹)及暗黑的球状体.发射星云和反射星云又称为亮星云,而亮星云和暗星云统称为弥漫星云.猎户座大星云是著名的弥漫星云.1888年,英国天文学家德雷耶(Dreyer,J. L. E.)刊布一册星团和星云的新总星表,简称NGC,公认为星云的标准星表.

气体星云(gaseous nebula) 一类星云.指主要由气体组成的亮星云.大部分分布在银道面附近,气体星云中也含有少量尘埃,其质量为太阳质量的几分之一到几千倍,大部分为太阳质量的10倍左右.著名的气体星云有猎户座星云、天鹅座网状星云和南天的古姆星云等.星云物质反射或散射附近的低温照明星的星光而发亮,具有吸收光谱的亮星云称为反射星云,如昴星团中的NGC1432星云.星云气体受其中或近旁的高热恒星的紫外辐射的激发而发光,具有发射光谱的亮星云称为发射星云.其中有高温核星、外形呈圆形或扁圆形的称为行星状星云.行星状星云的光谱包括连续谱和发射线两部分,连续谱分别由电子的自由-自由发射,受中央星加热的尘埃颗粒的热辐射以及氢和氦离子的复合过程.发射线大多是氢、氦、氮、氧的原子和离子的谱线,其中波长为5007Å和4959Å的谱线长期未能在实验室内找到,曾认为是由一种未知元素产生的,该元素定名为“氢”,以示是在星云内发现的.1927年,鲍恩(Bowen,I. S.)证认出这两条谱线为二次电离氧原子(OⅢ)所产生的禁线.由于星云内辐射能密度和电子数密度非常小,原子可以通过禁戒跃迁从亚稳态向低能态过渡,从而产生禁线.

反射星云(reflection nebula) 一种星云.指具有吸收光谱的亮星云.1912年,美国天文学家斯里弗(Slipher,V. M.)首先发现昴星团中的星云NGC1432具有吸收线光谱.星云物质反射和散射附近的低温照明星光而发亮,故名.照明星的光谱型都晚于B1型,温度较低,缺乏强烈的紫外辐射,不能激发星云,因而光谱中没有发射线.星云物质反照率很高,可能是些冰状粒子,由氢、碳、氮、氧的分子化合物组成,粒子的平均半径为1/4μm.

发射星云(emission nebula) 一种星云.指具有发射光谱的亮星云.形状不规则,因星云气体受其中或近旁的高热恒星的紫外辐射的激发而发光.1864年,英国天文学家哈根斯(Huggins,S.)首先发现此类星云光谱有发射线,表示它们由气体组成.除炽热气体(占总质量99%)外,星云中包含少量尘埃,发射星云物质粒子数密度为10^2～10^4个/厘米^3.电子温度一般在10000K左右.大部分分布在银河系的银道面附近和旋臂上.现已发现1200多个发射星云.猎户座大星云(M42)是典型的发射星云.另一著名的星云是人马座三叶星云(M20).

暗星云(dark nebula) 一种星云.指银河系中不发光的气体尘埃云.附近没有亮星或高温星的照射因而不发亮,只有当它们遮蔽了背景星光才显出形状.18世纪英国天文学家赫歇耳(Herschel,J. F. W.)发现了第一个暗星云,当时曾认为是天空中的空洞.暗星云中尘埃含量比亮星云略多.尘埃粒子的大小为10^{-7}m.银河系中暗星云很多,著名的暗星云有猎户座马头星云、南十字座中的“煤袋”星云、天鹅座的暗星云等.

行星状星云(planetary nebula) 一种星云.指发射星云中的一种.18世纪,英国天文学家赫歇耳(Herschel,J. F. W.)用望远镜观测这些星云呈圆形或扁圆形类似行星的视圆面,因而得名.已发现1500多个,角直径一般不超过几十(角)秒,其距离最近的为70～100光年,其余的均在几百光年以上.质量大约为太阳质量的0.2～0.6倍.行星状星云由气体和尘埃组成,星云物质粒子的数密度一般为10^8～10^{10}个/米^3,温度为1000～10000K.星云中心有一高温(50000K)核星,多为O型星或沃尔夫-拉叶星.星云气体由于核星的紫外辐射的激发而发光.

蟹状星云(crab nebula) 一个星云.指金牛座中的一个气体星云.1731年,英国业余天文学家比维斯(Bevis,J.)首次发现,1850年,英国天文学家罗斯(Ross,F. E.)据其形状取名为蟹状星云.1921年,比较相隔数年拍摄的照片,发现星云在膨胀,据膨胀速率推算,它是1054年超新星爆发后的遗迹,当时视星等估计为-5等,相应的绝对星等为-18等,与地球距离6300光年,角大小为$6' \times 4'$,线大小为12光年×7光年.它在射电、红外、紫外、X射线

以及γ射线波段也有强烈辐射. 观测证实其辐射是偏振的. 表明辐射源是相对论性电子在磁场中加速所产生的同步加速辐射. 它是惟一被认为具有同步加速辐射的超新星遗迹. 1968 年发现蟹状星云中核星为一脉冲星(NP0532 即 PSR0531+21), 脉冲周期为 0.0331s. 脉冲星的光学对应天体是一个亮度为 17 等的暗星. 1969 年后, 发现它还是光学脉冲星、X 射线脉冲星、γ射线脉冲星和红外脉冲星, 是一有极强磁场的快速自转的中子星, 直径约 10km. 据认为, 蟹状星云的能量来源于中子星的自转及脉冲星的电磁辐射.

猎户座星云(Orion nebula)　一个星云. 指猎户座中一个明亮弥漫气体星云, 又名 M42 或 NGC1976. 位于猎户座三星(猎户腰带三颗星 ζ, ε, δ)下方(猎户佩剑三颗星 σ, θ, ι)的 θ 星附近. 视星等为 4 等, 与地球距离约 1600 光年, 直径约 17 光年, 质量约 300 太阳质量. 星云中有很多高温的 O 型星和 B 型星, 它们发出紫外辐射使星云受激而发光, 是典型的发射星云, 其最亮部分原子数密度为 3×10^{10}个/米^3, 温度为 10000K, 其边缘温度约 6000K, 粒子数密度为 3×10^8个/米^3. 在猎户星云中发现了一些年轻恒星, 还发现了红外星、红外星云、球状体, 它们被认为是处于引力收缩阶段的恒星的前身. 射电观测表明, 它被一个半径为 200 光年、质量为 10^5 太阳质量的中性氢膨胀壳层所围绕. 它是一个比较年轻的天体, 年龄少于 100 万年.

赫比格-阿罗天体(Herbig-Haro object)　简称 HH 天体. 一类特殊天体. 为由红外源发出的星风激波与星际云相互作用产生的亮区. 由美国天文学家赫比格(Herbig, G. H.)于 1948 年和墨西哥天文学家阿罗(Haro, G.)于 1950 年发现, 故名. 由单个或多个星云凝块组成, 有以巴耳末发射线、电离硫和电离氧等禁线为主的低激发光谱. 位于银道面附近的暗星云内或其附近, 常伴有金牛座 T 型星. 现经光谱观测确认的赫比格-阿罗天体约 100 个.

球状体(globule)　一种星云. 指银河系中呈球状的浓密暗星云. 1946 年, 在一些亮弥漫星云背景上发现一些小的球状黑斑, 因而得名. 它们直径为 1000 到 10 万天文单位, 质量估计为 0.1～1000 太阳质量. 物质密度与体积介于恒星和星云之间. 温度约为 7～15K. 据认为它们是正处于引力收缩阶段的原恒星. 据估计, 银河系中较大的球状体大约有 20000 多个.

分子云(molecular cloud)　一种天体. 指银河系中某些星际分子集结成团而形成的天体. 其大小不一, 半径介于 10 天文单位和 33 光年之间. 大部分分布于银河系旋臂中. 由各种气态分子和尘埃粒子组成, 主要化学分子有羟基(OH)、一氧化碳(CO)、次甲基(CH)、氰基(CN)、氨(NH_3)、水(H_2O)等. 分子云密度相差很大, 分子数密度介于 10^8～10^{10}个/米^3 范围内, 其中心分子数密度可达 10^{12}个/米^3. 质量介于 10^4～10^7 太阳质量. 分子云通常是暗的, 用光学方法不易发现. 1963 年, 开始用射电方法观测星际分子 OH 的辐射. 1968 年, 在人马座 B2 星云观测到氨、水、甲醛(H_2CO)分子. 至今已发现分子云中有几十种星际分子.

星际分子(interstellar molecule)　一种天体. 指星际空间中的分子. 1937 年, 最早用光学望远镜发现了 CH(次甲基)、CN(氰基)和 CH^+(次甲基离子)的紫外吸收谱线. 1963 年用射电望远镜发现大量星际分子. 20 世纪 80 年代初已发现 64 种, 其中 46 种为由碳、氢、氧、氮组成的有机分子; 18 种为由硫、硅等组成的无机分子. 大多数为稳定化合物, 能在地球上找到; 少数在地球上根本不存在. 到 1994 年为止, 发现和证认的星际分子总数已达 108 种. 此外, 还有大约 50 种同位素分子以及 50 种左右已观测到其谱线发射而尚未证认出的分子.

电离氢区(ionized hydrogen region of galaxy)　简称 H Ⅱ 区. 一种特殊形式的天体. 指星际空间中, 星际气体氢大部分处于电离状态的区域. 在 O 型星、B 型星或由它们组成的星团的周围, 高温恒星发射出的紫外辐射, 使氢原子电离, 形成电离氢区. H Ⅱ 区中电离气体的典型温度为 8000～10000K, 离子数密度在 10^7～10^{12}个/米^3 范围内. 瑞典天文学家斯特龙根(Strömgren, B. G. D.)指出, 若取星际氢原子的平均数密度为 10^6 个/米^3, 则一颗高温的 O5 型恒星周围的 H Ⅱ 区的半径可达 450 光年; 对 B0 型恒星, 其周围的 H Ⅱ 区的半径为 160 光年; A0 型恒星周围的 H Ⅱ 区的半径只有 5 光年. 行星状星云和发射星云都是 H Ⅱ 区. H Ⅱ 区中的尘埃粒子吸收了高温恒星及电离气体的辐射后被加热到 20～300K 的温度, 因此几乎所有的 H Ⅱ 区都是红外源, 红外光度可达 10^7 倍太阳光度. 观测分析其光谱中的氢发射线, 或观测其射电波段中由电离氢的自由电子和质子复合过程产生的发射线均可研究 H Ⅱ 区.

星际磁场(interstellar magnetic field)　一种磁场. 星际空间中存在的磁场称为星际磁场. 其强度约为 10^{-10}～10^{-9}T. 分布不均匀, 有的地方较强, 有的地方甚至没有.

中性氢区(neutral hydrogen region)　简称H Ⅰ 区. 一种特殊形式的天体. 在星际空间中大部分以氢原子形式存在的星际气体区域. 大部分气体星云都处于中性氢状态, 距 O, B 型高温恒星 20～30 光年以外的区域, 由于能使氢电离的紫外辐射迅速减少, 星际气体也能够保持为中性氢区. 中性氢区的温度低于 100K, 平均为 60K, 原子数密度为 10^{-6}～10^{-5}

个/米³. 1944 年,荷兰天文学家范·德·赫斯特(Van de Hulst,H. C.)从理论上说明中性氢原子能够发出波长为 21cm 的射电辐射. 1951 年,美国射电天文学家珀塞尔(Purcell,E. M.)等人首次观测到来自银河系的 21cm 谱线的射电信号. 观测表明,中性氢区分布在银道面附近.

恒星天文学(stellar astronomy) 天文学的一个分支学科. 是主要以统计方法研究恒星和各种恒星集团的分布和运动特性的学科,同时研究取得基本数据的技术和方法,它是观测和理论相结合的一门科学. 恒星物理学着重恒星个体的研究,恒星天文学则着重于恒星集团的研究.

太阳运动(solar motion) 亦称太阳本动. 天文学术语. 指太阳相对于邻近恒星的运动. 1783 年,英国天文学家赫歇耳(Herschel,J. F. W.)分析天狼、北河二、大角等 7 颗恒星的自行,指出太阳朝向武仙座方向运动. 1837 年,德国天文学家阿格兰德(Argelander,F. W. A.)根据 390 个恒星的自行资料,证实了太阳运动的存在. 观测表明,太阳对邻近恒星的运动速度是 19.7km/s,朝向武仙座中赤经 $\alpha=18\text{h}4\text{min}\pm7\text{min}$,赤纬 $\delta=+30^\circ\pm1^\circ$ (1950.0 历元)的一点,该点称为太阳向点,简称向点;向点的银道坐标为银经 $L=56^\circ$,银纬 $B=23^\circ$. 在天球上与之相对的一点,称为太阳背点,简称背点,位于天鸽座内. 相对于 G,K,M 型暗星,太阳运动速度和向点与相对邻近恒星的有所不同. 运动速度为 35km/s,向点银道坐标为 $L=40^\circ,B=14^\circ$;相对于天琴座 RR 型星,运动速度为 140km/s,$L=88^\circ,B=12^\circ$.

平均视差(mean parallax) 亦称统计视差. 测定天体距离经常用的一种视差. 通过对一群恒星的运动资料进行统计分析而得到这一群恒星的视差的平均值称为平均视差. 它在恒星天文学中有广泛的应用. 进行三角视差的相对测量时,造父视差法中用于周光关系的零点的确定;分光视差法中,为测定分光绝对星等而绘出归算曲线时也需利用平均视差. 在其他视差不能很好测定的情况下,平均视差也可以给出单个恒星与太阳的距离. 1920 年,荷兰天文学家卡普坦(Kapteyn,J. C.)和范莱因(Van Rhijn,P. J.)通过对大量恒星的统计分析,得到视差 π、自行 μ 和视星等 m 的经验关系:

$$\lg\pi=-0.690-0.0713m+0.645\lg\mu.$$

统计视差(statistical parallax) 即"平均视差".

星团视差(cluster parallax) 天文学术语. 测定星团距离用的视差之一. 根据移动星团成员星的自行和视向速度的观测数据测定出的恒星视差称为星团视差. 若以 μ 和 V_r 分别表示星团成员星的自行和视向速度,θ 表示成员星与移动星团的汇聚点的角距(由观测得出),则成员星的视差为

$$\pi=\frac{4.74\mu}{V_r\tan\theta}\text{ 或 }\pi=\frac{4.74\mu}{V\sin\theta},$$

V 为整个星团的空间速度.

星团(cluster) 恒星的集团. 几十甚至上百万个恒星因恒星间的引力而聚集在一起的恒星集团称为星团. 按形态和成员星的数目可分为球状星团和疏散星团两类,前者因形状呈球状对称而得名,是由几万到几百万颗密集在一起的年老恒星组成,分布在银河系的外层空间(银晕)中;后者结构疏散,大部分位于银河系的中心平面(银道面),特别是银河系旋臂中,因而又称为银河星团,由几十到几百颗年轻恒星组成. 1784 年,法国天文学家梅西耶(Messier,C.)编制了包括 105 个云雾状天体的表(简称 M 表),其中有 57 个是星团. 除几个星团有专名外,星团常用 M 表的序号来命名. 1888 年后,英国天文学家德雷耶(Dreyer,J. L. E.)编制了《星团星云新总表》(简称 NGC),及其补编(简称 IC),星团也用 NGC 表或 IC 表的序号来命名.

疏散星团(open cluster) 一种星团. 指几十到上千颗恒星组成的彼此有物理联系的恒星集团,形状不规则,结构松散. 很久以前用肉眼就已发现了 4 个,金牛座的昴星团和毕星团、巨蟹座的蜂巢星团(鬼星团)及英仙座的双星团均为疏散星团. 现在发现的疏散星团数目超过 1000 个,估计银河系中疏散星团总数为几万个. 它们分布在银道面附近,离银道面的距离小于 650 光年,故又称银河星团,属星族 Ⅰ,大多数银河星团距离太阳在 1000 光年以内. 疏散星团的直径大多在 3～33 光年的范围内,成员星年龄为几百万年甚至几十亿年. 除主序星外,还有各种变星:金牛座 T 型变星、耀星、经典造父变星、Be 星、天鹅座 P 型星、Ap 星以及大量双星. 有些疏散星团与星云在一起,表明它们是年轻的.

球状星团(globular cluster) 一种星团. 指由几千到上百万颗恒星组成的恒星集团,密集成对称的球状. 用望远镜观测,在星团的中央部分恒星非常密集,不能区分开. 1665 年首次在人马座发现,至今已发现 132 个. 据估算,银河系中约有 500 个球状星团. 全天最亮的球状星团为半人马座 ω 星团,北天最亮的球状星团为武仙座的 M13 星团. 球状星团的直径在 16 光年到 350 光年. 分布在以银河系为球心的大致为球形的空间中,属星族 Ⅱ. 距银河系中心的距离大多数在 6600 光年以内. 它们在巨大的椭圆轨道上绕银河系中心旋转,轨道面和银道面的倾角很大. 有些星系际球状星团与地球距离大于 16500 光年. 球状星团成员星为老年恒星,年龄为 10^{10}a,有红巨星、天琴座 RR 型变星等. 天琴座 RR 型变星的光度彼此相差很小,平均绝对星等为 0.6 等,利用视星等 m、绝对星等 M

和距离 r 之间的关系式

$$M=m-5-5\lg r,$$

可由实际测出的视星等 m 求出距离. 对于发现其中有天琴座 RR 型变星的球状星团，便可利用此式来确定距离，对于未发现其中有这类变星的球状星团，可利用其中的红巨星(绝对星等为 -3 等)来确定它与地球的距离.

昴星团(Pleiades) 最著名的疏散星团之一. 位于金牛星座，其亮星构成二十八宿的昴宿，因而得名. 目力好的人可见七颗亮星，因用古希腊神话中阿特拉斯的七个女儿的名字命名，因而又称七姊妹星团. 成员星约有 280 颗，星团大小约 13 光年，与地球距离 417 光年. 星团中包含大量发亮的星云物质. 昴星团年龄约为 5×10^7a. 在古代，清晨看到它和太阳一起升起，定为航海和农耕季节的开始(春季)，清晨看到它没入地平线，标志着农耕季节结束(秋季).

毕星团(Hyades) 一个星团. 是位于金牛星座的一个疏散星团. 它的几颗亮星构成二十八宿的毕宿(毕宿五不属于毕星团)，因而得名. 与太阳距离 150 光年，大小约 33 光年，有 300 多颗成员星. 它还是个移动星团. 目前，毕星团正以 44km/s 的速度远离太阳. 毕星团年龄约为 4×10^9a.

移动星团(moving cluster) 一种星团. 其成员星以大致相同的速度和方向移动的星团称为移动星团. 已发现有 7 个移动星团，即毕星团、昴星团、大熊星团、鬼星团、英仙星团、天蝎-半人马星团、后发星团. 沿着或背着成员星的自行方向延伸，星团成员星大体汇合于一点. 前者称为汇聚点，后者称为辐射点. 有汇聚点的移动星团正在远离太阳，有辐射点的移动星团正在接近太阳. 毕星团是最早发现的移动星团. 1910 年，美国天文学家博斯(Boss, L.)发现毕星团的成员星汇聚于赤经为 6h17.5min，赤纬为 7°41′的一点. 毕星团离太阳 150 光年，大约有 300 颗恒星，向中心密集，较密集部分的角大小为 7°，线大小为 33 光年. 它在以 44km/s 的速度远离太阳. 大熊星团是最近的移动星团，由大约 100 颗恒星组成，北斗七星中的 5 颗、天狼星和南河三都属于大熊座移动星团.

星协(stellar association) 恒星集团的一种. 成员星具有相似物理特性的恒星集团称为星协. 其空间密度比星团稀疏，其中恒星彼此在不断分离. 1947 年，苏联天文学家阿姆巴楚米扬(Амбарцумян, В. А.)首先发现并定名. 星协分布在很大空间范围内，其成员星的空间密度甚至比周围恒星的空间密度还小，大致呈球状分布，集中在银道面附近. 分为两种：其一是 O 星协. 其直径在 30～200pc 的范围内，由十几至几百颗光谱为 O 型和 B 型的热蓝巨星组成. 有的 O 星协中也有沃尔夫-拉叶星、天鹅座 P 型星、B 型发射星和 M 型超巨星. 通常有一个或几个含 O 型星的疏散星团或四边形聚星或几颗星排成一线的星链作核心，已发现 O 星协约 50 个. 其二是 T 星协. 由金牛座 T 型星和御夫座 RW 型星组成，直径从几秒差距到几十秒差距，成员星数目从 10 颗到几百颗. 一般同星云在一起，已发现约 30 个. 1966 年，范登贝尔格(Van den Bergh, S.)提出反射星云内恒星聚集成 R 星协，直径为几个秒差距到 400pc，其成员星为 B0～A0 主型星，也有 A～M 超巨星. 据对星协成员星的自行和视向速度的分析，发现星协在扩张，说明星协是很年轻的不稳定的恒星系统，它们在 10^7a 内就会瓦解.

银河(Milky Way) 亦称天河、银汉、星河等. 一种天文现象. 指横跨星空的白色光带，它是银河系主体在天球上的投影，是由大量恒星、星云和星际物质构成. 银河在天鹰座与天赤道相交，依次经过天箭、海豚、狐狸、天鹅、蝎虎、仙王、仙后、英仙、御夫、金牛、双子、猎户、小犬等星座，穿过天赤道向南经过麒麟、大犬、船尾、船底、船帆、飞鱼、南十字、苍蝇、圆规、半人马、南三角、矩尺、天坛、南冕、天蝎、人马、盾牌、蛇夫、巨蛇等星座回到天鹰座. 银河的中线大致为天球上的一个大圆. 在天鹅座向南分成两支，西支经天鹅、天鹰，汇入蛇夫座；东支经狐狸、海豚、天箭、盾牌，汇入人马座. 从天鹅，经天鹰、人马、天蝎，到船底座这一段银河十分明亮. 银河各部分宽窄不一，从 5°到 30°，平均为 20°.

银河系(Galaxy) 一个星系. 指太阳所在的星系. 人们从银河系内所看到的它在天球上的投影，呈现为横跨天空的白色光带，称为银河，银河系因而得名. 银河系是由约 1000 亿颗恒星和大量星际气体及尘埃组成的旋涡星系. 其质量为 1.4×10^{11} 太阳质量. 银河系物质的主要部分集中在称为银盘的圆盘内，银盘中心隆起呈球状，称为核球，核球区域内恒星高度密集. 核球中心有个很小的致密区，称为银核. 银盘外面是一个范围广大、近似球状分布的系统，称银晕，其物质密度比银盘低得多. 银晕外面是银冕，大致是球状分布. 银盘中心平面称银道面. 综合光学和射电方法探测结果得知，银盘呈旋涡结构，恒星和星云物质大部分集中在四条旋臂中(即英仙旋臂、猎户旋臂、人马旋臂、三千秒差距旋臂)，银河系属 Sb 型旋涡星系. 银河系的不同区域里有不同类型的恒星. 星族 Ⅰ 的恒星分布在旋臂上，星族 Ⅱ 的恒星则分布在银晕及核球中. 银盘的直径约 75000 光年，太阳位于距银河系中心约 28000 光年，距银道面以北约 25 光年处. 银河系整体作较差自转一周的时间约 2.5 亿年. 银河系的年龄约 100 多亿年.

银心(galactic center) 天文学术语. 指银河系中心. 银河系自转轴与银道面的交点，位于人马座方向. 历元 2000.0 年的坐标为：赤经 17h45.6min，赤纬

—29°00′. 太阳距银心约 28000 光年. 银心为银道坐标系的起算点,其旧银道坐标为银经 327°41′,银纬—1°24′. 因星际物质的吸光作用,用光学方法很难看到银心. 综合射电、红外、X 射线、γ 射线探测结果得知,在离银心几十光年的范围内,为密集的恒星群及分立的气体尘埃星云. 恒星密度比太阳附近大 10^7 倍,估计在离银心 10 光年之内恒星总质量为 10^7 太阳质量,大约有 3×10^6 太阳质量密集在中央,推测银心区有一个大质量的黑洞.

银盘(galactic disk)　天文学术语. 指银河系的主体. 形如扁球的圆盘,中间厚,边缘薄,直径约为 75000 光年,中间厚度约为 6500 光年. 银盘以轴对称形式包围着银核. 主要由恒星组成,银盘的质量为 8×10^{10}太阳质量. 银河系有四条旋臂:人马旋臂、猎户旋臂、英仙旋臂和三千秒差距旋臂,它们都从银盘伸出,旋臂间距约 5000 光年,是气体、尘埃和 O,B 型星聚集的场所. 银盘外围被银晕包围着. 太阳是银盘内的一颗普通恒星,位于猎户座旋臂的内侧,以约 250km/s 的速度绕银心旋转,公转一周约 2.5 亿年. 星际气体散布在银盘内;星际尘埃大都集中在银道面附近.

银核(galactic nucleus)　天文学术语. 指银河系核球中的核心部分. 直径为 5 光年. 银核中的星高度密集,所拥有的质量相当于一千万个太阳质量,其中恒星的总质量约占几百万个太阳质量. 因在银道面两旁充斥着大量的星际尘埃,因此,在地球上用光学望远镜观测很难看清银核的结构. 由射电望远镜观测表明,银核是复杂的,中心有 5 个射电源,射电最强的一个可能就是银心. 银核又具有环状结构,发出强的射电、红外和 X 射线的辐射,物质状态还不清楚.

旋臂(spiral arm)　天文学术语. 指旋涡星系内年轻亮星、亮星云、经典造父变星等天体分布成旋涡状的条带. 是旋涡星系的主要外形特征,大多数旋涡星系有两条旋臂,少数的旋涡星系有 3 条以上. 旋臂内含有 O 型、B 型星及电离氢区,利用这些天体可确定银河系旋臂的位置. 此外,旋臂也是星际气体集结的场所. 观测中性氢 21cm 射电谱线也能探测银河系旋臂,而且射电观测不受星际尘埃的阻碍. 综合光学和射电探测结果,银河系有 4 条旋臂,它们为:英仙旋臂、人马旋臂、猎户旋臂和三千秒差距旋臂. 旋臂不是由同一些物质组成的"物质臂",而是其中的恒星有进有出,川流不息,旋臂中的恒星密度比旋臂外大(参见"密度波理论").

密度波理论(density wave theory)　天文学理论假说之一. 是说明星系旋涡结构的一种理论. 1942 年,首先由瑞典科学家林德布拉德(Lindblad,B.)提出,1964 年,由美国科学家林家翘建立较完整的理论. 该理论认为,恒星在绕中心旋转时,旋转的速度和空间密度都是波动变化的. 恒星的空间密度与旋转速度有关:绕转慢,恒星密集,反之,则稀疏. 由于密度像波浪一样变化,密度波既绕中心环行传播,又沿半径方向传播,因而密度极大的波峰呈现旋涡状分布,形成了旋臂. 旋臂中恒星密集,引力场加强,得以使旋臂自行维持. 旋臂中的恒星有进有出,川流不息,使得旋臂图案保持不变,不会缠绕起来. 密度波理论成功地说明了旋臂的形成和维持的原因,克服了从前认为旋臂是由物质构成的物质臂理论所具有的困难(按照这一理论,旋臂要么越来越紧,缠绕在一起,要么越来越松,旋涡图案消失).

银晕(galactic halo)　天文学术语. 指笼罩着银盘的巨大的球形稀疏雾状物. 直径约十几万光年(3×10^4pc). 由老年恒星和星系物质组成. 旋涡星系通常都有这样的晕,称为星系晕. 银晕包围着银盘,它的外侧又有银冕. 有人认为,银晕是一个旋转椭球体,长短轴之比为 2∶1,长轴为 5×10^4pc. 银晕中没有年轻的 O,B 型星和电离氢区,但均匀地分布着波长 3.7m 的射电辐射源. 银晕的估算质量为银盘质量的 10%,因对银晕的研究还不多,估算质量的差异可能甚大.

银河系次系(galactic subsystem)　一种恒星系统. 指银河系下一层次的恒星系统. 次系的概念首先由瑞典天文学家林德布拉德(Lindblad,B.)于 20 世纪 20 年代提出. 每个次系的成员在空间分布,运动特性和物理特性相近;不同次系则特性互有区别. 银河系由互相嵌套在一起的许多次系构成. O 型星次系、B 型星次系、经典造父变星次系和银河星团次系高度集中在银道面两侧,形成银河系扁平次系. 天琴座 RR 型变星次系、球状星团次系、亚矮星次系、室女座 W 造父变星次系,它们的银面聚度小,银心聚度大,形成银河系球状次系. 由蒭藁型长周期变星次系、新星次系、白矮星次系和 S 型星次系等组成的、介于银河系扁平次系和球状次系之间的称为银河系中介次系. 同类银河系次系的总和,称为银河系子系.

银河系子系(galactic component system)　一种天体系统. 指银河系下一次层次的天体系统. 银河系同类次系的总和(参见"银河系次系"). 各个扁平次系构成一个扁平子系,各个球状次系构成一个球状子系,各个中介次系构成一个中介子系. 银河系由三个子系套叠而成. 子系以及次系都是和星族(参见"星族")平行的天体系统.

星族(stellar population)　一种恒星集群. 银河系及旋涡星系中一些在年龄、化学组成、空间分布和运动特性十分接近的恒星集群称为星族. 1944 年,美国天文学家巴德(Baade,W.H.W.)研究仙女座星系 M31 和三角座星系 M33 中的恒星时,发现其核心部分和外围部分的恒星性质不同. 巴德将恒星分为两类:即星族Ⅰ和星族Ⅱ. 星族Ⅰ的恒星集中在星系外

围的旋臂内，其最亮恒星为早型白色超巨星，星际物质较多. 星族Ⅱ的恒星集中在星系核心部分，最亮星为K型红橙色超巨星，星际物质少. 现代把银河系所有天体分为五个星族：晕星族（年龄最老，如球状星团、亚矮星、周期大于0.4d的天琴座RR型变星组成）；中介星族Ⅱ（由高速星、周期小于250d光谱型、早于M5型的长周期变星组成）；盘星族（包括行星状星云、新星和银河系核心的恒星以及光谱中出现较弱金属线的恒星）；中介星族Ⅰ（包括光谱中出现较强的金属线星和A型星）；旋臂星族（年龄最轻，约几亿年甚至几千万年，主要有处于旋臂中的O型星、B型星、超巨星、疏散星团和星际物质）. 各星族的年龄相差很大，从晕星族、中介星族Ⅱ、盘星族、中介星族Ⅰ到旋臂星族，年龄依次递减. 一般地，较老的星族所含重元素的百分比，比年轻星族所含的要低，这是由恒星演化过程所决定的. 前一代恒星演化晚期所抛射的重元素进入星际物质，因此，由这种星际物质形成的年轻恒星包含较多的重元素.

银河系自转(galactic rotation)　银河系的一种运动. 指银河系内的恒星、星云和其他星际物质作为整体，在绕银轴作较差的转动. 自转速度与距离银轴远近而异. 距离很近时，速度与距离成正比，接近于刚体转动；距离很远时，速度与距离的平方根成反比，接近于开普勒旋转. 在太阳处，银河系的自转角速度为0.0053″/a，线速度为250km/s，周期为2.5亿年. 对银河系自转的观测有两种方法：其一是射电方法. 观测中性氢21cm谱线的位移，求得中性氢云的视向速度，从而推得银河系自转速度. 其二是光学方法. 主要根据视向速度和自行等资料求得银河系自转速度.

奥尔特公式(Oort formula)　表征银河系自转对视向速度和自行影响的公式. 1927年，由荷兰天文学家奥尔特(Oort, J. H.)导出，故名. 设ΔV_r表示银河系自转所产生的视向速度，也就是银河系自转对视向速度所产生的影响；r表示恒星与太阳的距离；恒星的银经、银纬分别为l, b；$\Delta\mu_l$表示银河系自转对银经自行的影响. 奥尔特公式为

$$\Delta V_r = Ar\sin 2l\cos^2 b,$$

$$\Delta\mu_l\cos b = \frac{A}{4.74}\cos 2l\cos b + \frac{B}{4.74}\cos b,$$

式中A, B称为奥尔特常数. 1964年，国际天文学联合会采纳的最佳值为

$$A = +15\text{km}\cdot\text{s}^{-1}/\text{kpc} = +4.87\times10^{-16}/\text{s};$$

$$B = -10\text{km}\cdot\text{s}^{-1}/\text{kpc} = -3.25\times10^{-16}/\text{s}.$$

如果用每年(角)秒表示，$A = +0.0032''/\text{a}$，$B = -0.0021''/\text{a}$. A, B与太阳处的银河系自转角速度$\omega_0 = V_0/R_0$存在关系

$$B = A - \frac{V_0}{R_0},$$

式中V_0为太阳绕银河系中心的转动速度，R_0为太阳与银心的距离. 1964年，国际天文学联合会采用的最佳值为

$$R_0 = 10\text{kpc},\quad V_0 = 250\text{km/s}.$$

据1986年用射电干涉法测定的结果，R_0为7kpc.

星系天文学(extragalactic astronomy)　天文学的分支学科. 是研究星系和星系际空间的天文学分支学科. 主要研究星系的结构、运动、起源和演化，星系、星系群、星系团的空间分布、相互作用、星系成团现象以及宇宙大尺度结构内的物质分布和高能活动. 研究手段是天体物理方法和射电天文方法. 常用动力学方法和统计学方法处理观测结果. 星系天文学是现代宇宙学的研究基础，是天文学中最活跃的领域之一.

星系(galaxy)　一种天体系统. 指由几十亿至数千亿颗恒星、星际物质构成的庞大天体系统. 星系占据着几千光年到几十万光年的空间，质量一般为$10^9\sim10^{11}$个太阳质量，与地球距离约为$2\times10^5\sim1\times10^{11}$光年，亮于23等的星系总数约有10亿个以上. 按形态分类，星系可分为旋涡星系(记为S)、椭圆星系(记为E)、棒旋星系(记为SB)、透镜星系(记为SO)和不规则星系(Irr). 银河系是一个旋涡星系. 按辐射特性和星系核活动的特性区分，有射电星系、赛佛特星系、马卡良星系、致密星系、N星系、类星体、蝎虎座BL天体等.

旋涡星系(spiral galaxy)　一种星系. 指具有旋涡结构的星系. 1845年，英国天文学家罗斯(Rosse, W. P.)首先发现猎犬座星系M51具有旋涡结构. 1926年，美国天文学家哈勃(Hubble, E. P.)首次提出星系分类法，将星系按形态分成旋涡星系、椭圆星系和不规则星系. 旋涡星系中心区呈透镜状，周围有扁平的盘状结构. 从中心区隆起的核的两端延伸出若干螺旋状旋臂叠加在星系盘上. 旋涡星系可分为Sa型(中心区大、旋臂紧卷)、Sb型和Sc型(中心区小、旋臂松弛). 旋涡星系的直径范围为16000～160000光年，其质量为$10^9\sim10^{11}$太阳质量. 星族Ⅰ天体组成星系盘和旋臂，星族Ⅱ天体构成星系核、星系晕(即笼罩星系外围的结构稀疏的天体系统)，星系晕外面的稀薄气体包层称星系冕. 在已观测到的星系中，旋涡星系占77%，由于旋涡星系比较明亮，易于发现，因此上述比例未必代表各类星系的真实比例. 银河系为一Sb型旋涡星系. 与地球距离最近的旋涡星系为仙女座星系M31.

棒旋星系(barred spiral galaxy)　一种星系. 指一种有棒状结构贯穿星系核的旋涡星系. 棒旋星系可分为三类：正常棒旋星系、透镜型棒旋星系(SBO)和不规则棒旋星系. 正常棒旋星系的棒状结构明显，有

旋臂从棒端伸出;透镜型棒旋星系有棒状结构而无旋臂;不规则棒旋星系的棒状结构不通过星系中心. 正常棒旋星系按旋臂缠卷得由紧到松的程度分为 SBa, SBb, SBc 三种次型. 棒旋星系在质量、光度、光谱、成员天体的星族类型、气体和尘埃的分布、星系盘和星系晕的结构等,均与正常旋涡星系相似.

椭圆星系(elliptical galaxy) 一种星系. 是呈圆球形或椭球形的星系. 1926 年,美国天文学家哈勃(Hubble, E. P.)首先提出按星系形态分成椭圆星系、旋涡星系等不同类型. 椭圆星系中心区域明亮,边缘区域亮度渐减. 按星系的扁圆程度可分为 E0, E1, E2, …, E7 八个次型. E 后面的数字 n 由下式确定:

$$n = 10 \times \frac{(a-b)}{a},$$

式中 a 和 b 分别表示椭圆的半长轴和半短轴. E0 型星系是圆星系, E7 是最扁的星系. 椭圆星系的大小为 $3\times10^4 \sim 6.5\times10^5$ 光年. 质量差别很大,小的仅与球状星团相当,有上百万个太阳质量,大的可达 10 万亿个太阳质量. 椭圆星系是由星族 Ⅱ 天体构成,没有星族 Ⅰ 天体,没有或仅有少量的星际气体和星际物质. 约 20%的椭圆星系中有中性氢、电离气体或发射 X 射线的炽热气体. 在所观测到的星系中,椭圆星系约占 20%.

不规则星系(irregular galaxy) 一种星系. 指外形不规则、没有明显的核心和旋臂以及盘状结构的星系. 可分为两类: Irr Ⅰ 和 Irr Ⅱ. 前者颜色偏蓝,色指数 B−V 约为 0.35;后者颜色偏黄,色指数 B−V 约为 0.8 左右. 这两类星系形状不规则的形成原因是不同的,前者的不规则性是固有的;后者是由类如星系核爆发、星系间的互相作用等扰动而引起. 它们是矮星系,质量为太阳质量的 $10^8 \sim 10^9$ 倍,大小介于 6500~30 000 光年,是由星族 Ⅰ 恒星、电离氢区、气体和尘埃组成. 麦哲伦星云就是不规则星系. 在观测到的星系中,不规则星系仅占 3%.

红移(redshift) 一种天体现象. 指天体电磁波谱中谱线位置向长波方向位移的现象. 红移 Z 的定义为

$$Z = \frac{(\lambda - \lambda_0)}{\lambda_0},$$

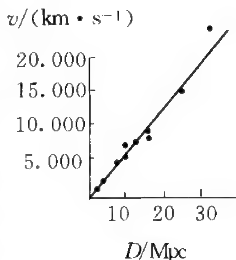
λ_0 是实验室光源的某一谱线的波长, λ 是天体的同一谱线波长. 根据多普勒效应,辐射源(天体)离观测者作退行运动时,谱线发生红移,趋近运动时则发生紫移. 红移 Z 与视向速度 v 成正比,可表示为 $Z=v/c$. 当视向速度 v 接近光速 c 时,红移

$$Z = \sqrt{\frac{c+v}{c-v}} - 1.$$

此外,根据广义相对论,处在引力场中的辐射源发生的辐射,对远离该引力场的观测者而言,谱线的波长也会发生红移,且红移量与辐射源同观测者之间的引力势之差成正比,称为引力红移. 假定星系的红移也是由多普勒效应引起,即与视向速度成正比,则根据哈勃定律,星系红移量越大,表示星系与地球距离越远,通常把它称为宇宙学红移. 如果认为星系红移与距离无关,则称为非宇宙学红移.

宇宙学红移(cosmological redshift) 见“红移”.

哈勃定律(Hubble law) 星系天文学的一个重要规律. 它描述河外星系的视向退行速度 v 和距离 D 的关系. 1929 年,美国天文家哈勃(Hubble, E. P.)发现,河外星系向四面八方的退行速度 v 与星系距离 D 成正比,即 $v=HD$,其中 H 为常数,称哈勃常数,单位为 km/(s · Mpc),一般取 $H=50$km/(s · Mpc). 由星系的红移大小可算出退行速度,进而利用哈勃定律可推算星系的距离. 哈勃和哈马逊(Humason, M. L.)得到的星系视向速度与距离的关系(如图所示).



哈勃常数(Hubble constant) 亦称速度-距离比. 一个著名常数. 指河外星系退行速度与星系距离的比值. 通常以 H 表示,单位为 km/(s · Mpc). 1929 年,美国天文学家哈勃(Hubble, E. P.)首先发现河外星系的视向退行速度与距离成比例,并给出比例常数为 500,称为哈勃常数. 1931 年,哈勃和哈马逊(Humason, M.)测定 H 为 526. 20 世纪 70 年代以来,用多种方法测定 H,采用 50~75.

互扰星系(interacting galaxy) 一类星系. 密近相遇或碰撞的星系对和多重星系称为互扰星系. 它们在引力作用下产生干扰,破坏了正常星系的形态而产生针状的、纤维状的或扫帚式的星系际桥或星系际尾. 在星系团内,星系之间的平均距离可小到 500kpc,星系的本动速度取为 1000km/s. 两个星系碰撞时间为 150 亿年,与星系的寿命相当. 两个星系迎头相撞,可能合并成为一个星系. 一些巨椭圆星系可能是吞食和它相撞的星系而逐步变大的.

活动星系(active galaxy) 亦称激变星系. 一类星系. 具有强烈活动的星系称为活动星系,它们具有小而亮的星系核,发出从射电到 X 射线的非热辐射,光谱中有宽的发射线,光度变化快速,有喷流结构. 成员星具有每秒几千公里的非圆周运动速度. 生存期短于普通星系. 赛佛特星系、马卡良星系、致密星系、N 星系等均属于活动星系.

激变星系(violent galaxy) 即“活动星系”.

特殊星系(peculiar galaxy) 一类星系. 形态和结构不同于正常星系的星系称为特殊星系. 其特征是

有亮的致密核，有很强的射电、X 射线辐射和红外辐射. 核的外围有物质抛射，核区光谱有强发射线和禁线，光度有快速变化. 有多种类型，如类星体、赛佛特星系、N 星系、射电星系、马卡良星系、致密星系、蝎虎座 BL 天体等.

赛佛特星系(Seyfert galaxy)　一类星系. 指具有强烈星系核活动的旋涡星系. 占全部已发现的旋涡星系的 1%，已发现 100 多个. 因美国天文学家赛佛特(Seyfert，C. K.)于 1943 年发现而得名. 它有一个明亮的恒星状核，光谱有很宽的发射线和禁线以及蓝连续谱，谱线有红移，光度很大，约相当于 10^{11} 太阳光度. 有强的红外、射电辐射，射电辐射具有非热的同步加速辐射性质，红外辐射大部分为星系核的光学辐射和紫外辐射加热的尘埃的再辐射，小部分为同步加速辐射. 光学波段的连续谱辐射主要是恒星热辐射. 亮度有变化，变化时标从 1 个月到 1 年. 现发现二十几个赛佛特星系有 X 射线辐射.

致密星系(compact galaxy)　亦称兹威基星系. 一类星系. 一种直径很小、亮度极大的星系称为致密星系. 由瑞士天文学家兹威基(Zwicky，F.)发现. 形状类似恒星，光谱红移极大. 其中红致密星系可能是表面亮度极高的正常星系；蓝致密星系可能是小尺度的星系际电离氢区. 它和类星体、N 星系、赛佛特星系等活动星系又通称致密天体.

N 星系(N galaxy)　一种特殊星系. 是周围有暗弱星云包层的致密星系，由美国天文学家摩根(Morgan，W. W.)于 20 世纪 50 年代提出. 其中心有高光度、恒星状的星系核. 发射强射电辐射. 光谱和赛佛特星系的类似，不过发射线较窄，星系核的亮度有变化.

马卡良星系(Markarian galaxy)　一类星系. 是具有反常强紫外连续谱的特殊星系. 因苏联天文学家马卡良(Маркарян，Б. Е.)发现而得名. 马卡良星系中有些是光谱中有宽发射线的赛佛特星系，有些是金属含量很低的不规则星系，有些是密近而有相互作用的双星系.

类星体(quasi-stellar objects or QSOs or Quasars)　一种特殊类型的天体. 指谱线具有巨大红移的特殊天体. 它们在天文照片上呈恒星状. 有强而宽的发射线，紫外辐射和红外光很强，一部分类星体射电和 X 射线辐射很强. 其直径仅为星系的十万分之一，但辐射能量却为正常星系能量的 200 倍. 现已发现 3000 多个. 大多数天文学家认为，类星体红移是宇宙学红移，遵从哈勃定律，因此是距离地球十分遥远的天体. 一般认为是遥远的活动星系核，其中心存在大质量的致密天体，可能是黑洞. 黑洞吸积着周围的物质，依靠引力能的转化，维持类星体的巨大辐射能量的输出.

蝎虎天体(BL Lac object)　河外天体的一种. 是一种具有亮中心核的椭圆星系，这类天体的原型是蝎虎座 BL，1929 年发现它有光变，曾当做变星，因而这类天体名称像一个变星名称. 其特征为：短时间曝光照片上呈恒星状，长时间曝光照片上有些可呈现出延伸状，即其周围有暗弱包层；光谱为连续谱，既无吸收线，又无发射线；射电、红外、可见光波段上的亮度都有快速变化，时标为几天到几个月；各波段上的辐射都是非热辐射，红外波段上辐射能量最多；辐射的偏振度大，且有快速变化. 现已发现 100 多个.

射电星系(radio galaxy)　一类星系. 是有很强射电辐射的星系. 射电功率为 10^{30}～10^{34}W 的星系称为正常射电星系. 射电功率比正常射电星系功率强 10^2～10^6 倍的星系称为特殊射电星系. 射电星系的射电结构有多种型式，它们大多是椭圆星系，很少有旋涡星系和不规则星系；有些是 N 星系和赛佛特星系. 很多射电星系也是强的 X 射线源和红外辐射源.

星系核(galactic nucleus)　天文学名词. 指星系的中心区域. 是恒星、电离气体、高能粒子的高度密集区. 正常星系的核是宁静的，可能含有各种致密天体. 星系核的大小约 100 光年，其质量为太阳质量的几百万到几千万倍. 少量的星系核有剧烈的活动，表现为剧烈的气体活动，气流速度达每秒几千公里；巨大的非热辐射在红外区达到极大，辐射功率达 10^{39}～10^{40}W；有强烈的光变和规模巨大的物质喷射等. 核活动最强的是类星体，其次是 N 星系和赛佛特星系.

仙女座星系(Andromeda galaxy)　一个著名的星系. 指仙女座中巨大旋涡星系(M31 或 NGC224). 历元 2000.0 年的坐标为赤经 0h42.7min，赤纬 41°17′，视星等为 3.5 等，肉眼可见状如椭圆云雾斑点，旧称仙女座星云. 距离地球 220 万光年，直径为 16 万光年，形状和结构类似银河系. 星系核由星族 Ⅱ 天体构成；星系盘则由星族 Ⅰ 天体构成，含有经典造父变星、新星、红巨星等天体；有九条旋臂，含有 O，B 型星、亮超巨星及 H Ⅱ 区；星系晕由球状星团组成. 总质量为 3.1×10^{11} 太阳质量. 它和银河系、大小麦哲伦云、三角座星系等约 40 个星系组成本星系群，它是其中最亮的，绝对星等为 −21.1 等. 它有两个伴星系(M32 和 NGC205)，近年发现稍远一些有 NGC147 和 NGC185 及三个更小的椭圆星系，它可能有七个伴星系.

麦哲伦云(Magellanic clouds)　银河系的两个伴星系. 指南天肉眼可见的云雾状天体. 其中较大的一个称大麦哲伦云，视角约 6°，位于剑鱼座和山案座之间；较小的称小麦哲伦云，视角约 2°，位于杜鹃座，合称麦哲伦云. 葡萄牙航海家麦哲伦(Magellan，F.)于 1521 年首次对它作了描述因而得名. 大麦哲伦云与地球距离 16 万光年，小麦哲伦云与地球距离为 19 万光年，是离地球最近的河外星系. 大麦哲伦云是棒

旋星系或不规则星系，质量为银河系的 1/20，约 1×10^{10}太阳质量；小麦哲伦云是不规则星系，质量为银河系的 1/100，约 2×10^{9} 太阳质量．它们和银河系有物理联系，构成一个三重星系．射电观测表明，它们有大量的星际气体，有一共同的很大的气体包层，且有中性氢气体流出．

本星系群(local group of galaxies) 银河系所在的星系团(参见“星系团”)．指以银河系为中心，半径为 300 万光年的空间内的星系集团．主要有银河系、大麦哲伦云、小麦哲伦云、仙女座旋涡星系 M31和 M32、三角座旋涡星系、NGC205，M33，NGC147，NGC185，IC1613 以及玉夫座星系、天炉座星系、NGC6822、狮子座 Ⅰ 星系、狮子座 Ⅱ 星系、天龙座星系、小熊座星系等 40 多个星系．总质量约为 6.5×10^{11}太阳质量，两个较大的星系——银河系和仙女座星系 M31 占本星系群总质量的大部分．本星系群内中性氢约占总质量的 1%．银河系相对于本星系群质量中心以 170km/s 的速度向银经 10°、银纬 10°方向运动．

星系团(cluster of galaxies or group of galaxies) 亦称星系群．天文学名词．指数目超过 100个的星系互相吸引聚集成的星系集团．成员星系数目较多的星系团称为富星系团，成员星系数目较少的称为贫星系团，成员星系数目不到 100 的星系团称为星系群．如后发座星系团，成员星系数多达数千个．星系团的大小一般为1500万光年左右．已知最远的星系团的距离在 70 亿光年以上，已观测到 1 万多个星系团．星系团有各种分类法，通常按其形态分成规则和不规则两类：前者成员星系呈球状分布，中心区星系高度密集，其成员大多为椭圆星系，总质量为 10^{15}太阳质量；后者结构松散，没有一定形状，比较大的不规则星系中有几个凝聚中心，其成员大多为旋涡星系和不规则星系，总质量在 $10^{12}\sim10^{14}$太阳质量．大多数星系团内部充满星系际热气体($10^{7}\sim10^{8}$K)，因而产生 X 射线辐射．

星系群(group of galaxies) 即“星系团”．

后发星系团(Coma cluster of galaxies) 一种星系团．指离地球较近的规则星系团，因位于后发座而得名．星系团角直径约 4°，可能包含大约 10000 个星系．据 1000 个星系的红移资料推算，其退行速度为6700km/s．距离为 4.5 亿光年．在中心附近有两个超巨星系：椭圆星系 NGC4889 和透镜星系 NGC1874．后发座星系团有延展的射电辐射和 X 射线辐射．

室女星系团(Virgo cluster of galaxies) 一个星系团．指与地球距离最近的不规则星系团，因位于室女座而得名．包含约 2500 个星系．成员星系中椭圆星系占 19%，旋涡星系占 68%．平均退行速度为1180 km/s，距离地球约 6000 万光年．超巨型椭圆星系 M87 位于星系团中心，它是全天最强的射电源之一，也是强 X 射线源．

超星系团(supercluster) 一种天体系统．指比星系团高一层次的天体系统．由若干星系团组成的集团，也称超星系，有的著作中称为星系云．一般拥有2～3个星团．其质量为 $10^{15}\sim10^{17}$太阳质量，具有扁长的形状，长径为 2～3 亿光年，短径为 5～7 千万光年．已发现由本超星系群、室女星系团、后发星系团等组成的本超星系团．较近的超星系团还有武仙座超星系团、北冕座超星系团等．

本超星系团(local supercluster) 一种超星系团．指包含本星系群在内的超星系团，也称本超星系．由本超星系群、室女星系团、后发星系团、大熊星系团和其他 50 个较小的星系群以及星系团组成的扁平状天体系统．直径约 1 亿光年，厚 500 万光年．在本超星系团中银河系距边缘约二三百万光年，绕本超星系团中心的公转周期约 1000 亿年．20 世纪 50 年代，美国天文学家德沃库勒(de Vaucouleurs，G．)分析了视星等亮于 13 等的 1000 多个星系的空间分布后，首先提出本超星系团的概念，并为后来的观测证实．

总星系(metagalaxy) 天文学名词．指现在观测所及的宇宙范围．根据目前观测水平，总星系的尺度为 200 亿光年，年龄为 200 亿年，包含大约 10 亿个以上的星系．总星系的物质分布是大体均匀和各向同性的．有种看法认为，总星系是比超星系团更高一级的天体系统．总星系所含物质中氢最多，其次是氦，1914年，美国天文学家哈勃(Hubble，E. P.)发现星系谱线有红移．如果把红移解释为多普勒效应，即星系退行所造成的结果，则表明总星系在均匀地膨胀．目前，流行的观点认为，造成总星系(或观测所及的宇宙)膨胀的原因是 200 亿年前一次大爆炸．

星系际物质(intergalactic matter) 一种天体．指星系际空间和星系团之间的空间中的物质．包括气体和尘埃，气体以氢为主，处于中性或电离状态．中性氢原子数密度不大于 10^{-9}个/米^{3}．“爱因斯坦 X 射线天文台”卫星揭示了电离氢气的分布情况．一种情况是成群地集结在星系或星系群周围，另一种情况是气体向星系团中心密集，密度向外平滑下降．

巨洞(giant void) 亦称星系短缺区．天文学名词．指在超星系团间存在的无星系的空区．观测表明，在 100 万秒差距的尺度上星系分布是不均匀的．在由富星系团密集在一起的成带状或片状区的超星系团之间就是星系短缺的空区，称巨洞．巨洞中缺少发光的星系，但存在着暗物质．

短缺质量(missing mass) 天文学术语．指星系团的位力质量与光度质量之差．根据位力定理算出的星系团总质量称为位力质量；而按照星系团的质光比估计的星系团总质量称为光度质量．前者约为后者的

10 倍，即与位力质量相比，质量有所短缺. 1933 年，瑞士天文学家兹威基(Zwicky,F.)首先发现后发星系团的质量短缺. 短缺质量是星系天文学中的一个重大课题，尚未得到满意的解决. 宇宙学中也有此问题，某些理论要求宇宙的质量密度为现在观测值的 40 倍，这是宇宙学中的短缺质量.

宇宙学(cosmology)　天文学的一个分支学科. 它从整体角度研究宇宙的大尺度结构、起源、运动和演化. 现代宇宙学包括观测宇宙学和理论宇宙学两个方面. 人们利用观测所及的时空范围(时间达百亿年以上，空间达百亿光年以上)内的观测结果研究宇宙的大尺度时空性质和物质分布、运动和相互作用的方式及规律，建立起宇宙学模型. 关于大尺度天体系统的结构，有两种不同的模型：一种是均匀模型，认为在大尺度上天体分布基本上是均匀的，并且在各方向上都是相同的. 另一种是等级模型，认为在任何尺度上，天体分布是不均匀的，呈逐级成团的分布. 关于河外天体系统性红移现象，有两种不同的理论：各种膨胀宇宙模型认为是系统性运动的结果，另外的理论是假定天体本身结构不同或光子传播过程中的能量衰减引起的. 关于大尺度特征变化与否也有两种不同的学说：稳恒态学说认为大尺度上的物质分布和物理性质不随时间变化；演化态学说则认为基本特征随时间有明显变化. 其中，又有主张温度是从高到低的热模型和主张温度是从低到高发展的冷模型两种. 在已有的各种宇宙学理论中以热大爆炸宇宙学最为成功. 它说明的观测事实最多，已成为标准宇宙学模型.

宇宙(universe or cosmos)　天文学名词. 指天地万物的总称. 天文学上特指人们观测到的各种天体及时空范围，常称为观测的宇宙或人们的宇宙. 随着科学的进步，观测的宇宙的范围不断扩大，从太阳系逐步扩展到银河系、河外星系、星系团，直到总星系. 现在已探测到 200 亿光年的宇宙深处，探索到 100 多亿年前的宇宙早期事件，已发现为数 10 亿以上的星系.

奥伯斯佯谬(Olbers paradox)　一种天文学悖论. 若宇宙中存在无限多个均匀分布的不动的恒星，则整个天空将是无限明亮，但实际夜空却是黑的. 这种矛盾是德国天文学家奥伯斯(Olbers,H. W. M.)于 1826 年提出的，故名. 现在一般用宇宙膨胀来解释，即膨胀宇宙中遥远天体的红移接近无限大，所发出的光子到达观测者时能量接近于零，因此夜空是黑的.

西利格佯谬(Seeliger paradox)　亦称西利格引力佯谬. 一种天文学悖论. 若宇宙中存在无限多个均匀分布的不动的恒星，则全部恒星对宇宙间任何质点(例如地球)的吸引力将是无限大或是一个不定值，但实际上地球受到的引力并非如此，理论与观测之间的这种矛盾，称为西利格佯谬. 这一矛盾是德国天文学家冯·西利格(Seeliger,H. von)于 1894 年提出的，故名. 1940 年，苏联天文学家埃根松(Эйгенсон,M. C.)证明，只有在天体均匀分布的宇宙里，引力矛盾才起作用. 在任何其他宇宙模型里，只要有一定的结构，西利格佯谬就不存在. 真实宇宙可能是有结构的.

宇宙学原理(cosmological principle)　天文学的一种理论假设. 指研究可观测的宇宙的大尺度时空特征时所采用的假设. 认为在宇宙的尺度上，在任何时刻物质的空间分布是均匀的、各向同性的，在空间任一点及任一方向物理量是完全相同的，但在不同时刻物理量可以不同，宇宙中各处的观测者所观察到的物理量和物理规律是完全相同的.

红移-视星等关系(redshift-apparent magnitude relation)　天文学术语. 这里指天体的红移与视星等之间的关系. 以天体的红移为纵坐标，视星等为横坐标，可以画出红移-视星等图，通常称为哈勃图. 由图可得出直接来自观测而未加任何假设的红移-视星等关系. 由此关系，1931 年，哈勃(Hubble,E. P.)假定宇宙间同类天体的绝对星等相同而且没有演化效应(即绝对星等不随时间变化)，肯定了视向速度 v 与视星等 m 之间的线性关系：

$$\lg v = 0.2m + B,$$

式中 B 为与绝对星等有关的常数. 只有当 m 的系数为 0.2 时，视向速度与视星等关系的线性才对应于视向速度与距离之间的线性. 1975 年，有人得到 1000 多个正常星系、射电星系和类星体的红移-视星等关系，三类天体的视星等值组合后得出系数为 0.3088. 测定红移-视星等关系和进一步确定视向速度-视星等关系，目的是得到天体的视向速度-距离关系.

速度-距离关系(velocity-distance relation)　天文学术语. 这里指星系的退行速度与距离间的关系. 美国天文学家哈勃(Hubble,E. P.)于 1929 年发现此线性关系. 这一关系不能直接测定，天文观测只能测定星系的红移与视星等的关系，必须作一些假定才得出速度-距离关系，而且只有红移较小时才存在线性关系；在红移较大时，要以相对论公式来代替多普勒效应公式，速度-距离关系就变得复杂.

哈勃距离(Hubble distance)　天文学术语. 指光在哈勃时间(哈勃常数的倒数)内所走过的距离，等于 c/H (c 为光速，H 为哈勃常数). 若取 H 等于 $55\text{km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1}$，则哈勃距离为 55 亿秒差距，即 180 亿光年，约为 1.7×10^{28}cm. 此距离亦即假定红移与距离的哈勃定律有效的距离.

宇宙半径(radius of the universe)　天文学术语. 在均匀各向同性的宇宙模型中，罗伯逊-沃克度规对椭圆空间、平直空间和双曲空间的空间曲率分别取为 +1，0 和 −1；曲率为 +1 时的椭圆空间的距离标

度因子称为宇宙半径(平直空间是无限的,双曲空间是开放的,故无半径).因只是对距离标度因子的几何意义的一种象征性解释,故常加引号,称为"宇宙半径".

宇宙年龄(age of the universe) 天文学术语.指宇宙模型中从某一特定时刻到现在的时间间隔.在宇宙常数等于零的宇宙模型中,宇宙年龄即代表宇宙现在的年龄.在宇宙常数小于零的各种宇宙模型中,宇宙年龄都比哈勃时间(哈勃常数的倒数,约为180亿年)小.在勒梅特宇宙模型中,宇宙年龄可大于哈勃时间.在稳恒态宇宙模型中,因为无论在什么时间都相同,所以不存在宇宙年龄.

宇宙常数(cosmic constant) 天文学术语.1917年,爱因斯坦(Einstein,A.)为了解释均匀和各向同性的静态宇宙的存在,在他的场方程中引进一个用符号Λ表示的项,在银河系范围内该项可忽略不计,只在宇宙量度中才可能有意义,故名.Λ可取作正数、负数或零值.1929年,哈勃(Hubble,E. P.)发现星系红移,确定静态宇宙模型不符合实际,爱因斯坦多次提出应取消Λ项,但不少学者认为Λ项可能有新的物理意义.例如,如果Λ为正,它表示宇宙中的一种斥力,两个星系距离愈大,Λ的值愈大,则它们之间的斥力也就愈大.估计宇宙常数的上限为$10^{-57}cm^{-2}$.有的学者认为Λ项为常数,有的则认为它随时间而变化.

爱因斯坦宇宙模型(Einstein cosmological model) 一种关于宇宙的理论假说.指爱因斯坦(Einstein,A.)于1917年建立的一个有物质无运动的有限无边的静态封闭宇宙模型.爱因斯坦在建立广义相对论后,将广义相对论的引力场方程用于整个宇宙,欲求一个均匀的各向同性的静态解,可是从引力场方程得不到这种解,他人为地加入一个宇宙常数项,它起着斥力的作用,从而得到一个非真空球状静态解.河外星系退行发现以后,这一模型即被否定.但它是现代宇宙学的第一个宇宙模型,被认为是现代宇宙学的开端.

弗里德曼宇宙模型(Friedmann cosmological model) 一种关于宇宙的理论假说.认为宇宙是均匀的、各向同性的、不断膨胀的、不含宇宙常数的一种宇宙学说,是1922年苏联数学家弗里德曼(Фридман,A.)首先求解爱因斯坦引力场方程得到的不含宇宙常数的各向同性的宇宙模型,故名.弗里德曼宇宙模型分为三种类型,分别对应空间曲率指数$K=1,0,-1$.$K=1$对应于三维球面空间,是一个有限的封闭宇宙,宇宙的膨胀将停止并转为收缩;$K=-1$和$K=0$分别对应于三维双曲面空间和平直空间,是开放的无限宇宙,宇宙膨胀将无限地继续下去.星系退行规律发现以后,这一模型受到重视,成为后来发展起来的一些宇宙模型的基础.

大爆炸宇宙论(big-bang cosmology) 一种关于宇宙的理论假说.认为宇宙是从密到稀、从热到冷不断地膨胀的一种宇宙学说.因为此演化过程好像一次巨大的爆炸,故名.为现代宇宙学中影响最大的一种学说.与其他宇宙模型相比,它能说明较多的观测事实,于1948年由天文学家伽莫夫(Gamov,G.)等提出.后来得到天体年龄、河外天体谱线红移、氦丰度和微波背景辐射等观测结果的支持,但对星系起源和各向同性分布等问题,这一学说尚未能解决.

膨胀宇宙模型(expanding cosmological model) 一种宇宙学模型.认为宇宙是不断膨胀的宇宙模型.已提出的属于这种模型的有弗里德曼宇宙模型、稳恒宇宙模型和大爆炸宇宙模型等.

暴胀宇宙模型(inflationary cosmological model) 一种宇宙学模型.指关于宇宙极早期的一种宇宙模型.1979～1981年由美国古斯(Guth,A. H.)、温伯格(Weinberg,S.)等人提出.该假说认为,宇宙在头一秒钟内从一个小的、均匀的、超密的"种子"以指数形式迅速暴胀.这种模型可以说明大爆炸宇宙模型难以解释的一些宇宙初始现象.

开放宇宙模型(open cosmological model) 一种宇宙学模型.指宇宙将永远膨胀下去的宇宙模型.它是弗里德曼宇宙模型的一个解,对应于三维双曲空间或平直空间,因而在此种宇宙中,光线永远回归不到"原处",即为开放宇宙模型.

减速因子(deceleration parameter) 宇宙学术语.为膨胀宇宙模型中的参量,以符号q表示.减速因子的现在值用q_0表示.在弗里德曼宇宙模型中,

$q_0>1/2$,对应于正曲率封闭空间;

$q_0=1/2$,对应于零曲率开放空间;

$0\leqslant q_0\leqslant 1/2$,对应于负曲率开放空间.

实测星系的数目、红移、视星等、视角和计算宇宙平均密度可得出减速因子.已得的结果差异很大,且互有矛盾,还不能明确确定.

稳恒态宇宙论(steady-state cosmology) 一种关于宇宙的理论假说.认为宇宙虽非静止,但在空间上是均匀的和各向同性的,在时间上是稳定的学说.1948年,由英国天文学家邦迪(Bondi,H.)、霍伊尔(Hoyle,F.)和戈尔德(Gold,T.)提出.这是一个单调膨胀的宇宙模型,要求在宇宙膨胀过程中物质密度不变,物质就必须连续不断地从虚空中创生,平均每5000亿年在$1m^3$内产生一个氢原子.这与已知物理学守恒定律和观测结果不符.

等级式宇宙论(hierarchic cosmology) 一种关于宇宙的理论假说.认为宇宙在结构上分成行星系、星系、星系团、超星系团等层次的学说.18世纪中期,德国物理学家朗伯(Lambert,J. H.)提出天体逐级成

团概念，1908年，瑞典天文学家沙利耶(Charlier,C. W. L.)提出等级式宇宙论，后来为美国天文学家德沃库勒(de Vaucouleurs,G.)所发展.这种宇宙论可解释奥伯斯佯谬，但因尚无精确的数学表述和理论预言，争议较多.

宇宙暗物质(invisible matter)　亦称不可视物质.一种特殊形态的物质.即宇宙中存在的不可视物质，观测表明在银河系周围的矮球星系的质量和光度比要比太阳的大30～50倍以上.这个结果证明，矮球星系中必定含有许多暗物质.矮球星系很少，要在这种小的星系中存在如此多的暗物质，则构成暗物质的粒子的质量必定较大.它们不可能是质量很小的电子中微子，而可能是其他的中微子.

天体演化学(cosmogony)　天体物理学的一个分支.研究各种天体的起源、发展和衰亡的学科.对太阳系的起源和演化，目前已有40多种学说，但尚无一种可被普遍接受的理论.目前较受重视的星云起源学说认为，太阳系是在约50亿年前由星际云中分裂出的一个原始太阳星云中形成的，原始星云有自转，在自吸引作用下收缩，中心部分形成太阳，外部形成星云盘，盘中的尘粒和小冰粒沉降到赤道面形成尘层，后聚集成固体块(星子).星子结合成行星和卫星.现代恒星演化学说认为，恒星是由稀薄的星际物质在自吸引作用下收缩形成的.收缩达到一定密度、温度和压力的时候，开始辐射能量并逐渐进行热核反应.恒星经历相当长的稳定时期(主序星阶段)后，为补充长期辐射所消耗的大量能量，内部开始收缩，收缩时释放的一部分能量使其外部膨胀，演变为低温度的红巨星，其后经历不稳定阶段抛射大量物质而最后演变为白矮星或中子星直到停止发光.对星系和星系团的形成目前还没有成熟的理论，对人们观测所及的宇宙的起源和演化属于宇宙学的范畴.宇宙热大爆炸学说较受重视.

宇宙化学(cosmochemistry)　天文学的一个分支学科.研究宇宙物质的化学组成及其演化规律的学科.是天文学和化学的交叉学科.依研究对象的层次和研究方法可以分为：陨石化学，研究各种陨石的化学组成；行星系化学，研究行星、卫星、小行星、流星体、彗星及行星际物质的组成和化学演化；恒星化学，研究恒星及太阳的化学组成和化学演化；星际化学，观测和发现星际分子，研究它们的形成和瓦解；同位素宇宙化学，确定宇宙物质中的同位素，研究元素的起源和演化；宇宙线化学，测定宇宙线中的介质和宇宙线源的化学组成.取得的主要成果是：测定出组成太阳等天体的化学元素及同位素的丰度；查明天体是由近百种化学元素及280多种同位素组成；发现了一些地球上不存在的矿物；发现了68种星际分子等.

射电天文学

射电天文学(radio astronomy)　天文学的分支学科.它是以无线电技术为观测手段观测天体在无线电波段的辐射(射电)来研究天文的学科.由于射电波长比光波长得多，它能通过光波通不过的星际尘雾，因而射电天文观测能够深入到光学观测看不到的区域.由于观测方法的发展和新技术的采用，射电望远镜系统的灵敏度已达微央斯基级.甚长基线射电干涉仪的分辨率可达万分之几(角)秒.射电天文学在天文学发展中做出重大贡献.例如，20世纪60年代的天文学四大发现：类星体、脉冲星、宇宙微波背景辐射和星际分子都是用射电天文手段发现的.

亮温度(brightness temperature)　天文学术语.描述天体亮度的一种物理量.指具有与天体亮度相同的绝对黑体的温度.常用以描述天体的射电亮度.很多天体射电是非热辐射，所以天体的亮温度与天体实际温度可能相差甚远.

央斯基(Jansky or Jy)　(1)央斯基(Jansky，K. G.)，美国无线电工程师的姓名.1932年1月，他首先在14.6m波长上探测到宇宙射电，开创了观测天体射电波的新纪元，诞生了射电天文学.(2)Jy，一种天位计量单位.天体射电流量密度的计算单位.等于$10^{-26}W\cdot m^{-2}\cdot Hz^{-1}$，用于量度宇宙射电流量密度.

宇宙射电(cosmic radio radiation)　天文学名词.是宇宙中除太阳系天体以外的各种天体发出的无线电辐射的统称，它包括银河系射电(中性氢区射电、电离氢区射电、星际非热射电、超新星遗迹射电、射电星射电、星际分子射电等)和河外射电等.

河外射电(extragalactic radio radiation)　天文学名词.是银河系以外各种天体发出的无线电辐射，包括：正常射电星系的射电，射电总功率约为10^{30}～10^{34}J/s；特殊射电星系的射电，射电总功率约为10^{32}～10^{40}J/s；类星射电源射电，射电总功率与特殊射电星系的相近；河外星系中性氢原子和星际分子(水、氨、一氧化碳等)射电以及微波背景辐射.

射电星(radio star)　一种射电源.指有连续谱射电辐射的角径小于1″的射电源.由于射电望远镜的灵敏度限制，除太阳外，目前尚不能观测到普通恒星的射电.所观测的射电星大部分具有特殊的射电爆发.可分为鲸鱼座UV型变星(红矮耀星)、红超巨星、射电新星、早型发射线星、射电双星、射电X射线星和脉冲星等七类.现已发现几千颗.

宁静太阳射电(quiet solar radio radiation)　天文学术语.指太阳射电中持久存在的一种背景辐射.从毫米波到10m波都存在，起源于太阳大气的热

辐射.

太阳缓变射电(slow varying component of solar radio radiation) 天文学术语.指太阳射电中缓慢变化的分量.与黑子和谱斑密切相关,存在于毫米波段到米波段,以分米波和厘米波段更为突出.起源于日冕凝聚区的热辐射.

太阳射电爆发(solar radio burst) 天文学术语.指太阳射电急剧变化的分量.爆发时间从1s到几天,流量可从宁静太阳的整个表面的射电辐射的百分之几到几十万倍.对于太阳射电爆发,目前还没有统一的分类方法.一般地,可分为微波爆发、分米波爆发和米波爆发.米波爆发来自日冕,又可分为噪暴(包括增强辐射和Ⅰ型爆发)、Ⅱ型爆发、Ⅲ型爆发及U型爆发、Ⅳ型爆发和Ⅴ型爆发.微波爆发来自色球——日冕过渡层,又可分为渐升渐降爆发、脉冲爆发和微波大爆发.分米波爆发分为分米波连续辐射和分米波快漂移爆发.

射电天文谱线(radio spectral line) 天文学名词.指天体的原子和分子在无线电波段的谱线.射电谱线的频带很窄,一般产生在低温、低压和低密度的星际空间.1951年,首次观测到中性氢原子的21cm的谱线;20世纪60年代首次测到电离氢区的复合谱线;到1985年已发现了66种星际分子的1800多条射电谱线,它们大部分集中在毫米波段.

宇宙射电源(cosmic radio source) 一种天体.有射电辐射的太阳系外的分立天体.已观测到3万多个,已证认为光学天体的只有几百个.可分为银河系射电源(包括超新星膨胀气壳、银河中性氢区和电离氢区、射电星、X射线源、分子云等)、河外射电源(有正常射电星系、特殊射电星系、类星射电源、河外星系中性氢和星际分子云)等.

类星射电源(quasi-stellar radio source) 一种宇宙射电源.指的是其光学对应天体外貌类似恒星、辐射谱线有很大红移的射电源,紫外辐射很强.其射电结构是具有两个射电子源,总能量很大,可达10^{55}J.现已发现500个以上的射电源.一般认为是星系演化早期的某一短暂阶段.

天体微波激射源(celestial maser) 一种特殊的星际分子.指一种微波射电很强、谱线很窄的特殊的星际分子(OH,H_2O,SiO…)源.多发生于原恒星区域和演化到晚期的长周期变量的星周包层,谱线辐射有的非常强.发出的射电在几周至几个月内剧烈变化,这种射电来自分子的受激发射放大.现已发现300多个星际分子源.

微波背景辐射(microwave background radiation) 天文学术语.指来自宇宙空间背景的各向同性的微波辐射.1965年,首先由美国科学家彭齐亚斯(Penzias,A.)和威尔森(Wilson,R. W.)在波长7.35cm上发现.其后许多研究者发现在从0.5mm至70cm波段内均有相当于温度为2.76K的黑体辐射的背景辐射,因此又称3K背景辐射.微波背景辐射被认为是宇宙原始大爆炸的残余热辐射.

空间天文学

空间天文学(space astronomy) 天文学的一个新的分支学科.采用高空飞机、高空气球、探空火箭、人造地球卫星、空间探测器、航天飞机等作为运载工具,在大气外对天体进行观测和研究.按观测的电磁波段,它开拓了红外天文学、紫外天文学,开辟了X射线天文学、γ射线天文学;按学科分支有空间天体测量学、空间天体物理学、空间射电天文学等.它突破地球大气的阻碍,可以观测来自天体的全波段电磁辐射;能够避开地球大气湍流对光线的影响,星像不被歪曲,从而提高了观测仪器的分辨能力.此外,还能够直接获取行星际物质粒子、月球表面物质样品和行星表面各种物理参量.空间天文学已取得许多重大成就.在近地空间探测中发现了地球辐射带、地球磁层;在行星际探测中发现了太阳风、太阳冕洞以及大行星的许多新卫星、环带、磁层和行星表面的多种物理参量;在红外探测中发现约25万个红外源等.又如,发现数千个X射线源、X射线谱线、X射线爆,以及γ射线爆等.这些重大发现对天文学产生重大影响,使人类对近地空间太阳系天体以及银河系、恒星、星际空间、星系际空间等的认识发生深刻变化.

红外天文学(infrared astronomy) 天文学的分支学科.指观测天体红外波段的电磁辐射来研究天体的学科.红外波段包括波长0.7～1000μm的范围,介于红光和微波之间.一般分为0.7～25μm的近红外区和25～1000μm的远红外区.除少数极狭窄波段(1.2,1.6,2.2,3.6,5.0,10.6和21μm)外,红外辐射受到地球大气(特别是水蒸汽、二氧化碳、臭氧)的严重吸收,因此大部分需利用高空飞机、高空气球或天文卫星进行观测.20世纪70年代以后,由于红外观测技术和空间科学技术的发展,红外天文学取得了巨大进步.除了设置地面大型红外望远镜外,1983年1月美、荷、英联合研制发射的红外天文卫星,连续工作10个月,共发现245 839个红外源(现已发现的红外源中,包括行星、彗星、小行星、卫星、恒星、新星、电离氢区、分子云、行星状星云、星系和类星体等).发现织女星周围可能存在行星系统.在红外波段还发现了新的星际分子谱线等.

红外源(infrared source) 一种天体.辐射能量集中在红外波段的天体称为红外源.已发现的红外源包括行星、卫星、彗星、早期和晚期恒星、行星状星云、电离氢区、分子云、银河系核、赛佛特星系、类星体等

各类天体.

BN 天体(BN object) 一种红外源.指猎户座中一个低温(530K)的红外天体.因贝克林(Becklin,E.E.)和诺伊格鲍尔(Neugebauer,G.)于1966年发现而得名.位于猎户座四边形聚星的北方约1′处.一般认为它是正形成的恒星,是原恒星的典型代表.

IRC 源(IRC source) 一种红外源.指美国加利福尼亚理工学院根据波长2.2μm的红外巡天观测所编的《红外源表》(Infrared Catalog,缩写为IRC)上的天体.辐射流量密度大约40Jy($1Jy=10^{-26}W\cdot m^{-2}\cdot Hz^{-1}$)的天体约5600个.

紫外天文学(ultra-violet astronomy) 天文学的分支学科.指观测天体的紫外波段的电磁辐射来研究宇宙的学科.紫外波段包括波长0.01~0.4μm的范围,介于可见光和X射线之间.一般把0.01~0.2μm称为远紫外区,0.2~0.4μm称为近紫外区.由于地球大气对紫外线的吸收,所以必须用高空气球及空间运载工具进行观测.在地球大气外可对太阳系天体进行整个紫外线波段的观测,对太阳系外天体,由于星际气体对紫外辐射的吸收,所以只能研究0.0912~0.3μm范围的辐射.紫外天文观测是研究F,G,K,M型等冷星的色球和恒星冕过渡区的重要手段,对热星来说,紫外天文观测可用来研究恒星风所造成的质量损失以及分析恒星的元素丰度.1978年1月发射的IUE(国际紫外探险者)卫星得到大量彗星、太阳系、冷星、热星、互扰双星、激变变星、星际物质、星系晕和活动星系的紫外辐射资料.

X 射线天文学(X-ray astronomy) 天文学的分支学科.指利用空间技术观测天体的X射线来研究宇宙的学科.X射线波段包括波长10^{-6}~10^{-2}μm的范围,介于紫外线和γ射线之间,一般可分为10^{-6}~10^{-4}μm的硬X射线和10^{-4}~10^{-2}μm的软X射线波段.天体的X射线受到地球大气的严重吸收,所以只能在150km以上的高空中进行观测.1962年发现了第一个X射线源(天蝎座X1)和来自银河系外的均匀的X射线背景辐射.1970年12月发射了第一颗X射线天文卫星后,又发射了一系列X射线天文卫星,已探测到几千个X射线源,其中有些已被证认为超新星遗迹、脉冲星、密近双星、X射线新星、宇宙X射线爆发和暂现X射线源、活动星系、星系团、类星体和宇宙X射线背景辐射等.对太阳的X射线观测发现,X射线冕洞就是高速太阳风的风源.

宇宙X 射线源(cosmic X-ray source) 一种天体.指辐射能量大部分集中在X射线波段的太阳系外天体.1962年首次探测到在天蝎座X1的X射线辐射,现已发现几千个X射线源.包括超新星遗迹、赛佛特星系、星系团等.在武仙座X1中还发现了58keV和110keV的X射线谱线.

宇宙X 射线暴(cosmic X-ray burst) 一种天体物理现象.指宇宙X射线在小于1s的时间内强度突然增强20~50倍的现象.总功率估计达10^{31}~10^{32}W以上.X射线暴源多数在银道面附近,少数在球状星团内.

γ 射线天文学(gamma-ray astronomy) 天文学的分支学科.指利用空间技术观测天体的γ射线来研究宇宙的一门学科.γ射线波段的波长短于10^{-6}μm,即短于X射线.由于地球大气对γ射线的严重吸收,所以只能在地球大气外进行观测.1962年发射的"月球轨道卫星徘徊者"3号和5号探测到宇宙γ射线背景辐射.1967年,"轨道太阳观测台"3号卫星探测到来自银盘的γ射线辐射."轨道太阳观测台"7号卫星探测到了和太阳耀斑相联系的γ射线爆发.现在已观测到某些脉冲星、活动星系核、暗星云等的γ射线辐射.γ射线源的数目达数十个.1967年,"维拉"卫星探测到持续时间几秒钟的宇宙γ射线爆发事件,其本质尚不清楚.

宇宙γ 射线源(cosmic γ-ray source) 一种天体.指辐射能量集中在γ射线波段的太阳系外的天体.1967年首次观测到来自银盘的γ射线辐射,现已在银河系内发现了几十个γ射线分立源,河外的γ射线分立源有类星体3C273和赛佛特星系NGC4151.

宇宙γ 射线暴(cosmic γ-ray burst) 一种天体物理现象.指能量为0.1~1.2MeV的软γ射线的强烈爆发,持续时间为0.1~100s,释放能量高达10^{33}J.人们分析美国监测核爆炸的"维拉"卫星观测资料首次发现.每年约检测到8次宇宙γ射线暴,暴源可能是银河系源.

撰 稿 马星垣 王绶琯 朱晔华 刘步林 刘麟仲
易照华 须同祺 贾沛璋 倪彩霞 阎林山
审 阅 马星垣 王绶琯 易照华

测　绘　学

测绘学(surveying and mapping)　研究地球和其他星体空间分布的有关信息的采集、测量、分析、显示、管理和利用的理论及技术的学科.研究内容包括:测定和推算地面点的几何位置、地球形状及地球重力场,获取地球表面自然形态和人工设施的有关信息,编制全球和局部地区各种比例尺的地图、专题地图,建立地理、土地等各种空间信息系统.

测绘学按其研究内容和应用,可分为普通测量学、大地测量学、实用天文学、摄影测量与遥感学、工程测量学、海洋测绘学、地图制图学等分支学科.测绘学的应用范围很广,它对国民经济建设、国防建设及科学研究等方面都起着极为重要的作用.

中国是世界文明古国,测绘方法的出现可以追溯到四千年以前.根据《史记·夏本纪》、《周礼·地官司徒》、《周礼·夏官司马》等史料记载,周代不仅有了地图和掌管地图的官职,而且那时地图已经有了各种用途.

春秋战国时期,测绘有了新的发展.从《周髀算经》、《九章算术》、《管子·地图篇》、《孙子兵法》、《孙膑兵法》等书的有关论述中知道,那时测量、计算技术以及军事地形图的内容和表现力已经达到了相当高的水平.

秦代曾修建郑国渠、灵渠、都江堰水利工程,并留下图书.根据史料记载,秦汉时期地图受到很大的重视,测绘有了进一步的发展.1973 年冬在长沙马王堆出土的汉代初期(公元前 2 世纪)长沙国的地形图、驻军图和城邑图,其内容的详细、方位的精确、设计的合理、符号的形象化、绘制的精美,都显示了中国当时测绘技术所达到的高水平.

西晋的裴秀主持编制了大型地图集《禹贡地域图十八篇》,并提出了分率、准望、道里、高下、方斜、迂直的"制图六体"作为制图准则,使中国测量和制图工作进入新的领域.唐代高僧一行(俗名张遂)于公元 724 年在河南平原主持了世界最早的子午线测量,比公元 814 年阿拉伯国家进行的子午线长度实测早 90 年.此外,唐相贾耽完成了《古今郡国道县四夷述》40 卷以及巨幅《海内华夷图》,在地图制图方面取得了成就.

宋元时期也出现了很多杰出的测绘成果.北宋沈括发展了裴秀的制图理论,发明了许多精密易行的测量技术,并且在世界上最早发现了磁针偏角.元代郭守敬在长期修渠治水实践中总结了一套水准测量的经验,首先提出了海拔高程的概念.朱思本用10 年时间编成了"计里画方"的《舆地图》.

明清两代,中国测绘技术又有了新的发展.明代郑和七使西洋,首次绘制了航海图.罗洪先改编《舆地图》为《广舆图》,明末祖绶又将其改制为《全国地图集》,即《皇明职方地图》.16、17 世纪,西欧测绘技术传入中国,与中国传统测绘技术相结合,中国开始出现了近代测绘.明万历年间,实测一些城市的经纬度,绘制了 10 多幅世界地图.清康熙年间进行了大规模的经纬度测量和地形测图,并在 1718 年编成著名的《皇舆全图》.清乾隆 21 年(1756 年)又补测了天山南北各地地图,并于 1761 年在《皇舆全图》的基础上补测汇编成《皇舆全览图》,覆盖范围较《皇舆全图》大一倍以上.

从 20 世纪初起,1903 年,清政府设立主管陆地测量的京师陆军测地局,各省设立分局,计划开展全国地形图的测图和编图.1904 年,又建立京师陆军测绘学堂,培养测绘人员.辛亥革命后,南京国民党政府设立陆地测量总局,各省分设陆地测量局,改组成立中央陆地测量学校.1912—1928 年,新测了部分地区1∶25000和 1∶50000 比例尺地形图,并于 1923 年至 1924 年,首次编绘出版了全国 1∶1000000 地形图 96 幅.1929 年,正式制订了大地测量规程,开展国家大地测量,并规定统一采用兰勃特投影.1931 年,陆地测量总局正式建立航测队,并协助水利、铁道、地政等部门组建航测队,测制了局部地区的军事要塞图、地形图、勘察设计用图,并进行了城市地籍测量.1932 年,在同济大学设置测量系,1945 年,中央测量学校也设立了大学部,为测绘事业培养了一些高级测量技术人才.从 1912 年到 1948 年,主要在中国东部地区进行了一、二等三角锁系和一、二等水准路线的基本大地测量工作,完成了约占全国总面积三分之一的 1∶50000 比例尺地形图,重编了全国 1∶1000000 比例尺地形图.在此期间虽然开展了一些测绘业务,但由于连年战乱,缺少完整的实施方案和全国统一的测量基准和技术标准,因此成果质量不高.

中国的测绘事业得到大规模发展始于 1949 年.在全国各省市和有关经济建设部门相继成立了测绘管理机构,并加强对测绘工作的规划和管理.在 1958—1963 年期间,先后制订了大地测量法式(草案)和各等级三角测量、基线测量、天文测量、水准测量、大地重力测量细则及各种比例尺地形图的测绘基本原则、图式、编绘规范,形成了一系列完整统一

的技术标准和技术规程,为提高基本测绘成果的质量提供了保证.大力发展测绘高等教育和中等教育,培养了大量测绘事业需要的高中级人才.建立了一系列专门的测绘研究所、研究室,一些测绘生产、教学、勘测部门也设置了测绘科学研究机构,全国形成了门类比较齐全的测绘科学研究体系.科研工作取得了丰硕成果,促进了测绘事业的发展和提高.

从 20 世纪 50 年代初迄今,中国的测绘工作已经取得了巨大的成就.随着一、二等大地测量工作的开展,陆续建立起包括大地原点、青岛水准基点在内的各种大地测量基准.1982 年已在全国布测天文大地网点近 5 万个,并完成天文大地网整体平差.1976 年以后,重新布测和完成全国一等水准测量 100 个环,共 9 万多公里.完成了包括重力基本点和一等点的国家重力控制网.除部分地区外,主要采用航测方法完成了全国基本地形图.20 世纪 70 年代中期,基本完成了覆盖全国大陆的 1∶50000 比例尺为主的地形图.20 世纪 80 年代初,完成了全国 1∶200000、1∶500000、1∶1000000 比例尺地形图的编制出版.完成了近海区域图测绘和大量海图集的编制,并逐步开展了海洋重力和某些远海测绘工作.测绘工作在各项工程建设中,包括厂矿、铁路、公路、水运、水利枢纽、城市市政、建筑、军事等工程建设中,都起到了"排头兵"的作用,做出了很大的贡献.

从 20 世纪 70 年代起至 90 年代,测绘技术有了重大突破.空间技术、各类对地观测卫星使人类有了对地球整体进行观察和测绘的工具,好像可以把地球摆在实验室进行观察研究一样方便.由空间技术和其他相关技术,如由计算机、信息、通信等技术发展起来的 3S 技术(即全球定位系统(GPS))、遥感技术(RS)和地理信息系统(GIS))在测绘学中的出现和应用,使测绘学从理论到手段都发生了根本性变化.

GPS 卫星定位的出现革新了传统的地面定位方式,大大提高了大地测量定位的效率和精度,并已应用到工程测量和航空摄影测量;卫星激光测距(SLR)技术和甚长基线干涉(VLBI)技术已成为监测地球自转运动变化和地壳运动的理想工具;各类卫星重力探测技术的迅速发展填补了全球海洋重力测量的空白,首次为人类提供了高分辨、高精度的地球重力场模型,给出了更精密的地球形状参数.大地测量学已迎来了空间大地测量时代,并发展到包括时间在内的四维大地测量.

传统的摄影测量数据主要来自卫星遥感或数字摄影获得的影像数据,测绘人员在室内借助高速、高容量计算机和专用配套设备对遥感或摄影影像的记录数据进行地表(甚至地壳浅层)几何或物理信息的提取和变换,得出数字化地理信息产品,由此制作各类可供社会使用的专用地图等测绘产品.地图制图学也由传统的纸质人工地图编制、生产和更新发展到计算机扫描数字化和自动或半自动戚图,形成各类地图数据库,中国 960 万平方公里国土的国家基本地图的成图或更新周期可望从十几年、几十年缩短到几年或更短.

今天,光缆通信、卫星通信、数字化多媒体网络技术可使测绘产品从单一纸质信息转变为磁盘和光盘等电子信息,产品分发可从单一邮路转到电路(数字通信和计算机网络传真),测绘产品的形式和服务社会的方式正在向信息化的方向发展.

中国的测绘学科和测绘事业在新技术的支持和推动下,近 30 年来有了很大的进步.20 世纪 80 年代初建立了卫星多普勒定位网,20 世纪 90 年代又建立了国家 GPSA 级和 B 级定位网,形成了新一代中国定位坐标框架;以监测地壳运动为主要目的的全国高密度点位的 GPS 监测网络也已初步建成;GPS、遥感技术、地理信息系统以及数字化测图、制图系统等新技术已开始普及并得到广泛应用.

普通测量学

普通测量学(elementary surveying)　测绘学的一个分支.即研究地球表面局部区域内测绘工作的基本理论、技术、方法和应用的学科.局部区域指在该区域内进行测量、计算和绘图时,可以不顾及地球曲率的影响,而把这区域的地面简单地当做平面处理.

普通测量学研究的主要内容有局部区域内的控制测量和地形图的测绘.测量工作包括距离测量、角度测量、高程测量以及观测数据的处理和地形图的测绘.控制测量是为测绘各种大比例尺地形图而建立的一定精度的平面控制网和高程控制网.地形图测绘是在控制测量的基础上,将地球表面的地物和地貌(总称为地形)按一定比例尺测绘成图.此外,应用普通测量学的技术、方法和仪器,还可进行一般工程的施工测量.

铅垂线(plumb line)　地球重力场中的力线.即重力的方向线,或悬挂重物时自由下垂的直线.

铅垂面(vertical plane)　亦称垂直面或竖直面.与重力场相关的一个平面,即包含测站铅垂线的平面.

垂直面(vertical plane)　即"铅垂面".

竖直面(vertical plane)　即"铅垂面".

平均海水面(mean sea-level)　描述海面状态的一个面.指验潮站用验潮仪在海边观测多年所得的海水水位升降的平均位置.它是高程起算的基准面.平均海水面在不同时期和不同地点略有差异.中

国以往曾有多种高程起算的水准零点，如沽口零点、吴淞口零点、废黄河零点、坎门零点等．为了统一，后来规定按1950—1956年青岛验潮站历年观测数据计算的黄海平均海水面作为全国高程起算的水准零点．由这一基准面起算的高程系统称为1956年黄海高程系统．现在又有新的1985年高程系统．

水准零点（height datum） 一种测量基准点．即国家高程控制网高程起算的零点（参见“平均海水面”）．

绝对高程（absolute elevation） 简称高程．亦称海拔或海拔高度．地面点空间位置的一个分量，即地面点沿铅垂线方向到大地水准面之间的距离．

海拔高度（height above sea-level） 即“绝对高程”．

海拔（height above sea-level） 即“绝对高程”．

相对高程（relative elevation） 亦称假定高程．表示地面点空间相对位置的一个分量，即地面点沿铅垂线到某一指定水准面的距离．当测区附近没有国家水准点时，可假定某一点的高程为起算值，以确定其他各点的高程．待与国家水准点联测后，即可换算成绝对高程．

假定高程（assumed elevation） 即“相对高程”．

水准测量（leveling） 亦称几何水准测量或抄平．一种测定高程的基本方法．即用水准仪和水准尺测定地面两点间高差的方法．在地面两点间安置水准仪，观测竖立在两点上的水准尺，利用水准仪的水平视线在尺上读数可推算两点间的高差．一般由水准原点或任一已知高程点出发，沿选定的水准路线依次测定高差，然后根据已知点的高程推算出沿线各水准点的高程．由于不同高程处的水准面互不平行，沿不同路线测得的两点高差将有差异，所以在整理国家水准测量成果时，需按所采用的高程系统加入必要的改正，以求正确的高程．

中国国家水准测量依精度不同分为一、二、三、四等．一、二等水准测量称为精密水准测量，可为研究地壳形变等提供数据．三、四等水准测量直接为地形测图和工程建设提供所必须的高程控制．

几何水准测量（geometric leveling） 即“水准测量”．

精密水准测量（precise leveling） 见“水准测量”．

抄平（leveling） 即“水准测量”．

水准点（benchmark） 测量控制点的一种．即用水准测量方法测定的高程控制点（参见“水准测量”）．

高差（height difference） 一种相对高程概念．即任意两点高程之差．若 A 点比 B 点低，则 A 点到 B 点的高差为正；反之，高差为负．高差通常用水准测量和三角高程测量的方法测定，也可用气压高程测量的方法测定，但后一种方法的精度远低于前两种方法．

比高（relative elevation） 一种相对高程概念．即地物与地貌相对于某一地面点的高度．在地形图上，常对陡岸、路堤、独立石、城墙、宝塔、梯田坎等注出比高．

水平面（horizontal plane） 一种局部测量基准面．即与水准面相切的平面，它与某切点的铅垂线正交．因为地球的半径很大，所以小范围内的水准面都可看成是水平面．

水平线（horizontal line） 一种局部测量基准方向线．即位于水平面内的直线，亦即与铅垂线正交的直线．

比例尺（scale） 表示地图详细程度的一种规格．即地图上一条线段的长度与地面上相应线段实际水平距离之比．直接用数字表示的称为数字比例尺，一般用分数 1∶M 表示，M 为分母，表示缩小为若干分之一．另有图解式的直线比例尺和复式比例尺（也称斜线比例尺）．通常在图上都绘注有比例尺．比例尺越大的地图描述的内容越详细．

平面图（planimetric map） 地图的一种．即不考虑地球曲率，把小块地区的地球表面当做平面所绘制的正射投影图．有时也把仅表示地物、不表示地貌的图称为平面图．是一种简略地图或工程应用地图．

地形图（topographic map） 地图的一种．即按一定的比例尺，表示地物、地貌平面位置和高程的正射投影图．地形图是经过实地测绘或根据实测和配合有关资料编制而成．

地形测量（topographic survey） 一种主要的测绘工作．它是为获得地形图而进行的测量工作．即为了按一定比例、按规定的图式符号和注记绘制成地形图，而对地球表面的地貌、地物在水平面上的投影位置和高程进行的测量工作．它包括图根控制测量和碎部测量，按“先控制后碎部”的程序进行作业．大面积地形图的测绘可采用摄影测量的方法；小面积或专用于某项工程建设的地形图，过去一般采用大平板仪或平板仪加经纬仪的白纸成图方法，现在逐渐推广应用计算机辅助的数字化测图．

地物（planimetric feature） 地图要表示的基本要素之一．指地球表面上相对固定的物体，如居民点、道路、河流、森林和各种工程建筑物等．一般分别用统一规定的符号表示．

地貌（relief） 地图要表示的基本要素之一．指由于地球内外应力的作用以及人类改造自然等因素，使地球表面呈现高低起伏的形态，如山地、丘陵、

平原、高原、凹地、冲沟、陡崖等.地貌一般用等高线表示,也有用晕渲、晕渲等方法表示.

地形(landform) 测绘工作要测定和描绘的主要对象.它是对地物和地貌的总称.

等高线(contour) 地图上表示地面高程的曲线.即地面上高程相等的相邻各点所连成的闭合曲线,用以显示地面的高低起伏形态,并有一定的立体感.地形图上的等高线主要有首曲线(基本等高线)和计曲线(加粗等高线).有时为显示首曲线不能显示的局部地貌特征,可加绘间曲线和助曲线.

等高距(contour interval) 地形图上表示高程变化的一种方式.即地形图上相邻两条等高线之间的高差.在同一幅地形图上,等高距是相同的.等高距越小,表示地形起伏变化越精细.

首曲线(intermediate contour) 亦称基本等高线.是按规定等高距描绘的等高线(参见"等高线").

基本等高线(basic contour) 即"首曲线".

计曲线(index contour) 一种作了标注的等高线.即为了判读和计数方便,每隔四条(或三条)基本等高线描绘的一条加粗等高线(参见"等高线").计曲线上注有高程,以便判读.

间曲线(half-interval contour) 亦称半距等高线.是一种附加等高线.即按二分之一等高距描绘的等高线,图上常以长虚线表示,一般用于显示平缓山顶、鞍部、微型地貌及倾斜变化的地形情况.

助曲线(supplementary contour) 亦称辅助等高线.是一种附加等高线.即按四分之一等高距描绘的等高线,图上常以短虚线表示,一般用于显示平坦地段的地面微小起伏情况.

示坡线(slope line) 一种地图符号.即等高线上指示斜坡降落方向的短线,它与等高线垂直正交.一般在山头、谷地和斜坡方向不易判读的地方以及凹地的最低和最高的等高线上绘出.

大地子午面(geodetic meridian plane) 与定义大地经度相关的平面(参见"大地经度").即包含某一地面点和参考椭球极点、极轴的平面.

磁子午面(magnetic meridian plane) 与地球磁场方向相关的平面.即过地面一点并包含地球南北磁极的平面.

子午线(meridian) 表示南北方向的一种平面曲线.即大地子午线与天文子午线的简称.大地子午面与参考椭球面的交线称为大地子午线,天文子午面与参考椭球面的交线称为天文子午线.

真子午线(true meridian) 与地球南北极相关的线.即通过地面某点及地球南极、北极的平面与地球表面的交线.习惯上将真子午线指向北极的方向作为真北方向.

磁子午线(magnetic meridian) 与地球磁场方向相关的线.即通过地面某点及地球磁南极、磁北极的平面与地球表面的交线.习惯上将磁针北端所指的方向作为磁北方向.

方位角(azimuth) 量度方向的一个角度.即从某点的基本方向线北端起,按顺时针方向量至目标方向线的水平夹角.其值自 0°到 360°.通常用真子午线、磁子午线和纵坐标轴方向作为基本方向.以真子午线方向为基本方向的为真方位角;以磁子午线方向为基本方向的为磁方位角;以纵坐标轴方向为基本方向的为坐标方位角.

真方位角(true azimuth) 见"方位角".

磁方位角(magnetic azimuth) 见"方位角".

坐标方位角(grid azimuth) 在平面坐标系中度量方向的角度.即从纵坐标轴指北方向起,顺时针量得某方向线的方向角.

方向角(bearing) 量度方向的角度.即从某特定方向起,按顺时针方向量得的某方向线的水平角.

磁偏角(magnetic declination) 与局部磁场方向相关的角度.即磁子午线与真子午线间的夹角.这是因为地磁两极与地球两极不重合所引起的.磁子午线北端偏于真子午线以东为东偏,其值为正;偏于西者为西偏,其值为负.磁偏角的大小和正负因地而异,同一地点的磁偏角也随时间而有微小变动.

象限角(quadrant angle) 量度和标定方向的一种角度.常用的有磁象限角和坐标象限角.前者是指测线与磁子午线所夹的锐角,后者是指平面坐标系中测线与纵坐标轴所夹之锐角.象限角由南或北读起,读为偏东或偏西若干度.象限角不大于 90°.

角度测量(measurement of angle) 测量的基本工作之一.指测定水平角或竖直角的工作.水平角是确定地面点的平面位置的要素之一,竖直角用以确定地面点的高程.

水平角(horizontal angle) 推算边长、方位角和点的坐标的主要元素.一点到两个目标的方向线垂直投影在水平面上所成的角.

垂直角(vertical angle) 亦称竖直角.推算两点间高程差的主要元素.是从一点至观测目标的方向线与一特定方向之间在同一竖直面内的夹角.通常以水平方向或天顶方向(铅垂线的反方向)作为特定方向.目标方向线和水平方向线(水平线)间的夹角称为高度角 (α),其值自 0°到±90°.目标在水平线以上的高度角为正,称为仰角;目标在水平线以下的高度角为负,称为俯角.目标方向线和天顶方向线间的夹角称为天顶距 (Z),其值自 0°到 180°.

竖直角(vertical angle) 即"垂直角".

高度角(altitude angle) 见"垂直角".

仰角(elevation angle) 见"垂直角".

俯角(depression angle) 见"垂直角".

天顶距(zenith distance)　见"垂直角".

坐标格网(coordinate grid)　地图的一个基本组成部分.即地图上用来确定点位坐标的格网.分为地理坐标格网和直角坐标格网两种.前者也称制图网,是以一定经度、纬度间隔按某种地图投影方法描绘的经线、纬线网,注有经度、纬度,便于确定点位的地理坐标;后者也称公里网,一般是平行于所选定的直角坐标轴,并按一定间隔描绘成正方形格网,注有公里数,用于确定点位的平面直角坐标.

直角坐标格网(rectangular coordinate grid)　亦称公里网.坐标格网的形式之一.是平行于直角坐标轴并以一定间隔表示在图上的方格线网,注有公里数,供展绘点位和确定点位的平面直角坐标之用(参见"坐标格网").

公里网(kilometeric grid)　即"直角坐标网".

地形图分幅(separate sheet of topographical map)　地形图的测制和使用的一种编号管理方式.为了统一规划测图工作和便于利用、保管地形图,而按一定规则划分图幅的方法.有经纬线分幅和直角坐标网矩形分幅两种.前者一般用于百万分之一至万分之一地形图的分幅,它是以经差6°、纬差4°的一幅百万分之一地图的图幅为基础,按一定的经差和纬差进行划分的.例如,按经差3°、纬差2°分为4幅50万分之一图幅,再按经差1°、纬差40′分为36幅20万分之一的图幅,又按经差30′、纬差20′分为144幅10万分之一图幅,再以十万分之一图幅为基础分为5万分之一、2.5万分之一和万分之一的图幅.大于五千分之一的地形图,用直角坐标网进行矩形分幅.两种图幅均用文字和数字按一定顺序进行编号.

真北(true north)　用于确定方位的一种基准方向.即地面点真子午线北端所指的方向,即真子午线指向地球北极的方向.通常用天文测量的方法测定,也可用陀螺经纬仪直接测定.

磁北(magnetic north)　用于确定方位的一种与磁场方向有关的基准方向.它是地面点磁子午线北端所指的方向,即磁针指向地磁北极的方向.通常用罗盘仪测定.

控制测量(control survey)　一种基础性测量.在一定的区域内为地形测图或其他目的建立控制网所进行的测量工作.控制网包括平面控制网和高程控制网.平面控制网与高程控制网一般是分别单独布设,有时也可布设成三维控制网.控制网具有控制全局、限制测量误差累积的作用,是各项测量工作的基础和测算依据.平面控制测量包括三角测量、导线测量、三边测量和边角测量;高程控制测量包括水准测量和三角高程测量等.

小三角测量(minortriangulation)　一种控制测量方法.即小测区建立平面控制网的一种方法.用于小测区的首级平面控制或三、四等三角网以下的加密,是建立图根平面控制的一种方法.根据测区情况可布设成单三角锁、中点多边形、大地四边形和线形三角锁等不同图形.其特点是:边长短,精度要求较低,计算时不考虑地球曲率的影响.

三边测量(trilateration)　建立大地控制网和工程测量控制网的一种方法.它是利用电磁波测距仪测定水平控制网(锁)中各三角形的三边长度,通过几何关系推算三角形各顶角的大小,从而计算出各边的方位角和三角形顶点的坐标.

边角测量(combination of triangulation and trilateration)　建立平面控制网的一种方法.在测角网(锁)中加测全部或部分边长,可以提高网点精度,但工作量较大.一般多用于精度要求较高的工程测量和形变监测的平面控制网中.

三角网(triangulation network)　布设水平控制网的一种形式.在三角测量、三边测量、边角测量中互相连结的三角形所构成的控制网.构成网状的称三角网,构成锁链状的称三角锁.

三角锁(triangulation chain)　见"三角网".

三角点(triangulation point)　组成三角网的基本元素.即三角网(锁)中各三角形的顶点.通常在地面上埋设标石,作为三角点实地位置,并根据实地情况,在点上设置不同类型的觇标,供观测使用.三角点的平面坐标、高程,以及与相邻三角点间的边长和方位角等,称为三角点的成果,是地形测量和工程测量的必要资料.

线形三角锁(linear chain of triangulation)　一种低等级控制测量方法.即在两个已知点间用连续三角形加密控制点的一种方法,因其构成的图形为线形而得名.线形三角锁选点方便,计算简单,加密边缩短迅速,但因校核条件少,一般只在直接用于控制测图的低等级控制测量中采用.

基线(baseline)　三角测量中的一种尺度控制.三角测量中推算三角网(锁)各边长度所依据的起算长度.其水平距离用因瓦基线尺、电磁波测距仪或其他精密测距仪等精确测定.有时在工程测量网中所依据的基本线,也称基线.

基线测量(baseline measurement)　一种尺度控制测量.即精确测定三角网(锁)中基线长度的工作.过去用因瓦基线尺丈量.20世纪50年代以来,由于电磁波测距仪的发展,目前大部分已用电磁波测距仪直接测定.

图根点(mapping control point)　测图控制测量专用语.即用图根测量方法测定的控制点.

碎部测量(detail survey)　地形测量中测绘地形图的一个工序.指在图根控制测量后,首先在测图

板上按坐标展绘各级控制点，然后根据控制点测定地物和地貌的特征点（如房角、道路交叉点、山顶、鞍部等）的平面位置和高程，对照实地将有关地物、地貌按比例尺以相应的符号描绘在图上．地物特征点和地貌特征点统称为碎部点．测定碎部点的平面位置，最常用的方法是极坐标法；碎部点的高程，通常用视距测量法测定．按所用仪器的不同，碎部测量常用的方法有：平板仪测图、小平板和经纬仪联合测图、经纬仪测绘法等．

碎部点（detail point） 见“碎部测量”．

交会法（intersection method） 根据已知点的平面位置测定未知点的平面位置的一种方法．分解析交会法和图解交会法．一般分为三种情况：

1．在两个已知控制点 A，B 上，分别对待定点 P 观测水平角，从而求出待定点平面位置的，称为前方交会法（图 1）．

2．由两个已知控制点 A，B 和待定点 P 组成的三角形中，在一个已知控制点（例如 A 点）和待定点 P 上分别观测水平角，而求出待定点平面位置的，称为侧方交会法（图 2）．

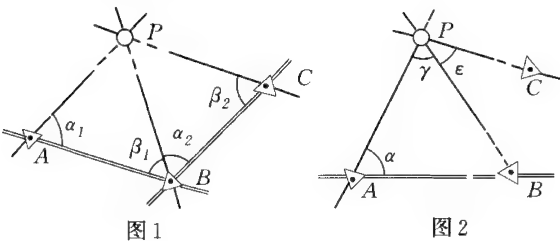


图 1　　图 2

3．在待定点 P 上分别对三个已知控制点 A，B，C 观测三个方向间的两个水平角，从而求出待定点平面位置的，称为后方交会法（图 3）．

图 3

为了检核和提高点位精度，前方交会应在三个已知控制点上进行交会，侧方交会一般在待定点上对第三个已知控制点再观测一个水平角，后方交会则需观测四个已知点方向间的三个水平角．

前方交会（forward intersection） 见“交会法”．

后方交会（resection） 见“交会法”．

侧方交会（side intersection） 见“交会法”．

平板仪测量（plane-table survey） 用平板仪扩展、图解图根点及测绘地形图的方法和过程．即将测区所需要的已知控制点按一定比例尺展绘在测图板上，并注明其高程；然后用图解法加密图根点；再以上述各点为测站，进行碎部测量，从而测绘成地形图．在平板仪测量中，水平角是在图板上直接图解．若待定点的平面位置用图解交会法测定，则点与点之间的距离也可从图上量得．

小平板仪测量（topographic mapping with alidade） 一种类似平板仪测量的方法．即用测斜仪代替平板仪所用的望远镜进行低精度碎部地形测量的一种方法．它的应用原理与平板仪相同，构造比平板仪简单，搬站容易，速度快．小平板仪常与经纬仪或水准仪配合使用进行测图．

视距测量（stadia survey） 一种常用测量方法．即用具有视距装置的测量仪器同时测定两点间水平距离和高差的方法．一般用经纬仪或平板仪望远镜内的视距丝（在十字丝分划板上与横丝平行的两根上下对称的短丝）照准标尺，读取标尺上的读数，并测出竖直角，即可算出标尺与仪器间的水平距离．若同时读取横丝在标尺上的读数，并量出仪器高，则还可求得两点间的高差．这种方法简便、迅速，不受地面起伏的限制，但精度较低．主要用于地形测量．

三角高程测量（trigonometric leveling） 一种低精度高程差测量方法．即通过观测两点间的水平距离和竖直角求出两点间高差的方法．也是测量地面点高程的一种方法．在测站点上测定至照准点的竖直角，量取测站点的仪器高和照准点的目标高，若已知两点间的水平距离，按三角形边角关系可以计算出两点间的高差，从而求出高程．当两点间距离较远时，计算高差时需加大气折光影响的蒙气差改正．由于观测竖直角受大气垂直折光的影响不易进行精确改正，所以三角高程测量的精度低于水准测量．但因其作业简单，在山区和丘陵地区仍得到广泛应用．

气压高程测量（barometric leveling） 另一种低精度高程差测量方法．即根据大气压力随高程变化的规律，用气压计进行高程测量的一种方法．由于大气压力受气象变化的影响较大，因此气压高程测量的精度远低于水准测量和三角高程测量的精度．主要用于低精度的高程测量．但它的优点是使用方便、经济、迅速．

距离测量（distance measurement） 测量的基本工作之一．它是测量地面上两点连线长度的工作．距离测量的精度用相对误差表示，即距离测量的误差同该距离长度的比值，用分子为 1 的分式表示．比值越小，距离测量的精度越高．距离测量常用的方法有钢尺（或皮尺）丈量法、视距法、视差法和电磁波测距法等．通常需要测定的是水平距离．

图根控制测量（topographic control survey） 亦称地形控制测量．地形测图的一个基础性工序．即是直接为地形测图建立平面控制和高程控制所进行的测量工作．一般是在高级控制点间逐步进行加密，以满足测图的需要．图根平面控制常采用小三角测

量、导线测量或交会法等方法;图根高程控制通常用水准测量或三角高程测量等方法.由图根控制测量测定了平面位置和高程的控制点称为图根点,也称地形控制点,是测绘地貌点、地物点的平面位置和高程的依据.

地形控制测量(topographic control survey) 即“图根控制测量”.

地形控制点(terrain control point) 见“图根控制测量”.

大地测量学

大地测量学(geodesy) 测绘学的一个主要分支.即研究建立地球空间定位坐标框架,在三维时变空间测定和描述地球的形状、运动状态、重力场及其随时间变化的学科.

建立地球空间定位坐标框架是:在大地坐标系或以地球质心为原点的空间直角坐标系中,建立高精度三维地面控制网,作为地面任何待定点定位的参考框架.为地图测绘、国家经济建设和国防建设服务是大地测量学的工程应用任务.

地球自然表面,即大陆地形面和不受干扰的海洋表面是地球的真实形状,通常用大地水准面的形状代表复杂的地球真实形状;又确定一个最佳逼近大地水准面的旋转椭球,代表一种简单规则的地球形状,这是大地测量学研究地球形状的任务,包括测定大地水准面的形状和确定相应的地球椭球的形状及大小.地球形状是地球形成和演化动力过程的结果,也与地球内部物质结构有关,因此研究地球形状对认识地球形成和演化过程以及内部物质结构有重要科学意义;同时确定地球椭球的形状和大小及其在地球内部的定位、定向又是建立大地坐标系的首要工作,与其工程应用任务密切相关.大地水准面是与全球平均海水面密合的一个重力等位面,需要进行全球重力测量确定它的形状.大地水准面是地球重力场在地球表面的几何表象,测定和描述地球重力场是大地测量学的主要任务之一.地球重力场的结构反映地球内部密度的分布,重力场随时间的变化,是地球系统物质迁移过程的反映.地球重力场数据是研究地球物质分布和运动以及现今全球变化的基础信息.任何地球外空飞行体的运动都受地球重力场的支配,精密地球重力场模型支持卫星的精密定轨,支持航天技术的发展.地球体及其内部物质在各种内力和外力作用下处于不断的运动状态,监测地球自转的变化、日月引潮力产生的潮汐现象,以及地球内部动力过程产生的板块运动和地壳形变等相关地球动力现象,是现代大地测量学的重要任务,为研究地球动力环境变化及预测相关自然灾害提供大地测量监测数据.上述是大地测量学的科学任务,与其工程应用任务又是相互支持和促进的.

现代大地测量学包括几何大地测量学、物理大地测量学和空间大地测量学三个主要部分.空间大地测量学是近30年来,在卫星定位等空间技术的支持下发展起来的一门大地测量学新分支,它标志着大地测量学的一次跨时代的技术革命.卫星定位技术已经普及,并用于监测地球的运动和变化;卫星重力探测技术正在迅速发展,由此,人类首次获得了精密地球重力场模型.现代大地测量已经由静态大地测量发展到动态大地测量,由区域大地测量发展到全球大地测量,从而扩大了大地测量学与相关地球学科交叉发展的广度和深度.

由于航天技术的发展,现代大地测量学也包括研究月球和行星表面的大地测量问题,并形成行星大地测量学或称宇宙大地测量学的新分支.

几何大地测量学(geometric geodesy) 亦称天文大地测量学.经典大地测量学的主要分支.是研究用几何法测定地球形状和大小以及地面点几何位置的学科.它采用一个同地球外形最为接近的旋转椭球代表地球形状,用天文大地测量方法测定该椭球的形状和大小,并以它的表面为参考面,研究和测定大地水准面,建立大地坐标系和推算地面点的几何位置.

地面点几何位置的测定包括水平控制测量和高程控制测量两部分.水平控制测量的基本方法有三角测量、边角测量和导线测量等;高程控制测量方法有水准测量和三角高程测量等.分别用于建立一个国家的水平和高程控制网.

天文大地测量学(astro-geodesy) 即“几何大地测量学”.

大地水准面(geoid) 根据地球重力场理论研究地球形状的一个基本概念.重力位等于常数的所有空间点形成的连续曲面称为重力等位面,通常称为水准面.处于静止状态的水面是一个水准面.大地水准面是无穷个水准面中的一个代表地球整体自然形状的特定水准面.李斯廷(Listing,J.B.)于1973年首先提出这一概念,设想处于静止平衡状态的海洋面向大陆下面延伸,所形成的封闭曲面称为大地水准面(这是大地水准面的经典定义).在大地测量中,用验潮站长期观测所确定的平均海面(代替理想的静止海面)定义大地水准面上的一个点,并作为海拔高程的零点,因此,大地水准面又称水准基面.现代海洋学结合卫星测高技术已证实,平均海水面并不是一个重力等位面.事实上,即使排除了日月引力的潮汐影响,由于海面大气压力、温度和盐度的分布差异以及海洋上的面流作用,使平均海面不是只受重力场作用的重力平衡的自由表面,这样一个表面

在物理上不是等位面.设想一个“最接近”平均海水面的重力等位面，平均海水面相对于这个等位面的高程称为海面地形，其值可达1～2m.在理论上，将对应的海面地形全球积分平均值为零的等位面作为大地水准面，这是大地水准面现代定义的基本涵义.但是由于在实践上根据全球海面地形确定一个能普遍用于海洋学和大地测量学的大地水准面还存在困难，因此，目前还没有一个公认的统一定义.

水准面(level surface)　见“大地水准面”.

水准基面(geoid)　见“大地水准面”.

参考椭球(reference ellipsoid)　用于建立大地坐标系的规则参考形体.根据一个国家或地区的大地、天文和重力测量资料推算出的一个与该区域的大地水准面最为接近的旋转椭球，以此作为处理本国或本地区大地测量成果的参考面.

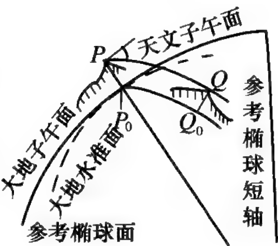
参考椭球的基本元素有长半轴 a、短半轴 b 和扁率 α ($\alpha = (a - b)/a$) 等，知道其中两个元素值便可把参考椭球的形状和大小确定下来，一般使用 a 和 α.中国自1952年起采用克拉索夫斯基参考椭球，其元素为

$$a = 6\ 378\ 245\mathrm{m}, \quad \alpha = 1 : 298.3.$$

1980年后改用国际大地测量学与地球物理学联合会第16届大会推荐的1975年椭球体元素：

$$a = 6\ 378\ 140\mathrm{m}, \quad \alpha = 1 : 298.257.$$

参考椭球定位(orientation of reference ellipsoid)　建立大地坐标系确定其参考曲面的基本步骤.即按一定条件将具有确定元素的地球椭球同大地体(大地水准面的封闭形体)的相对位置固定下来，由此确定大地测量计算基准面(参考面)和大地起算数据的过程.



参考椭球定位的一般方法是：在建立国家大地网初期，于选定的大地原点 P 上精密测定该点的天文经度 λ_0、纬度 φ_0 以及至一相邻点 Q 的天文方位角 α_0，并用水准测量测定该点的正高 H_0^g.然后把 λ_0 和 φ_0 分别作为大地经度 L_0 和纬度 B_0；把 α_0 作为大地方位角 A_0；再将 H_0^g 作为大地高，这相当于使 P 点处的大地水准面与参考椭球面的差距 $N'_0=0$.并假定参考椭球短轴与地球自转轴平行，起始大地子午面与起始天文子午面平行(见图).图中的 P_0 点为 P 点沿垂线方向在参考椭球面上的投影点，此时，P 点的垂线与 P_0 点的椭球面法线重合；图中的 Q_0 点为 Q 点沿垂线方向在参考椭球面的投影点，一般情况下，除 P 点假定其垂线与参考椭球法线重合外，其他点的垂线不一定与参考椭球法线重合.经过这一初始定位过程，参考椭球在大地体内的位置就完全固定下来了.这种定位方法称为“一点定位”.若国家大地网已建成，从大地原点出发可以推算其他天文点的大地经度和纬度，并与这些点的天文经度和纬度比较，便可求出各点的垂线偏差分量 ξ_i' 和 η_i'，进而可以求出这些点的 N_i'.在重新定位时，令

$$\xi_i = \xi_i' + \mathrm{d}\xi_i,$$
$$\eta_i = \eta_i' + \mathrm{d}\eta_i,$$
$$N_i = N_i' + \mathrm{d}N_i,$$

在 $\sum(\xi_i^2 + \eta_i^2) =$ 极小，或 $\sum N_i^2 =$ 极小的条件下，解得大地原点上修正后的 ξ_0，η_0 和 N_0，从而获得适合于该国家和地区的大地起算数据：

$$B_0 = \varphi_0 - \xi_0,$$
$$L_0 = \lambda_0 - \eta_0 \sec\ \varphi_0,$$
$$H_0 = H_0^g + N_0,$$
$$A_0 = \alpha_0 - \eta_0 \tan\ \varphi_0,$$

它们分别为大地原点的大地纬度、大地经度、大地高和大地方位角.这种定位称为多点定位.

多点定位(multi-point orientation of reference ellipsoid)　见“参考椭球定位”.

大地原点(geodetic origin)　亦称大地基准点.定义国家大地坐标系的一个基本概念.即国家水平控制网中推算大地坐标的起算点.在大地坐标系中，大地原点的大地经度、大地纬度、大地高和至另一相邻点方向的大地方位角，是推算国家水平控制网中各点大地坐标的依据，称为大地基准点数据，它定义和实现一个国家的大地坐标系.中国的大地原点设在陕西省西安市附近的泾阳县，根据该原点推算而得的坐标系，定名为1980年大地坐标系.

大地基准点(geodetic datum)　即“大地原点”.

大地基准点数据(geodetic datum data)　见“大地原点”.

大地坐标系(geodetic coordinate system)　亦称地理坐标系或椭球坐标系.是大地测量学中最常用的一种坐标系.大地坐标系的基本参考面是经过椭球定位后的参考椭球面.参考椭球的短轴平行于地球的协议平自转轴(通过协议国际地极原点CIO的自转轴)，参考椭球的起始大地子午面平行于格林尼治平均天文台子午面(由多个天文台的观测数据推算确定的格林尼治平均天文台(经度零点)，并包含协议平自转轴的格林尼治子午面)，这时参考椭球的赤道平面与地球平赤道面(过地球质心与协议平自转轴正交的平面)平行.地面一点在大地坐标系中由三个坐标分量确定其位置，即大地纬度 B、大地经度 L 和大地高 H，这三个坐标分量总称为大地坐标.大地纬度 B 是过该地面点的椭球面法线与赤道面相交而成的锐角(0°～90°)，北半球大地纬度为

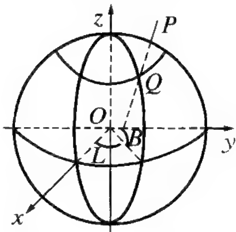
正，自赤道向北量取，称为北纬；南半球大地纬度为负，自赤道向南量取，称为南纬．大地经度 L 是过该点的大地子午面与起始大地子午面的夹角（0°～180°），从起始大地子午面向东量取为正，称为东经；向西量取为负，称为西经．大地高 H 是该点沿椭球面法线到参考椭球面的距离．(B,L) 是大地坐标的水平分量，H 是高程分量．大地坐标系分参心大地坐标系（又称区域大地坐标系）和地心大地坐标系（又称全球大地坐标系）两类：前者中参考椭球短轴不一定与地球协议平自转轴重合，但平行，参考椭球面只和一个区域大地水准面密合；后者中参考椭球又称总地球椭球，其短轴与地球协议平自转轴重合，且椭球中心与地球质心重合，椭球面与全球大地水准面密合．一个国家的大地测量常采用适合于本国的参心大地坐标系；全球大地测量和卫星大地测量均采用地心大地坐标系．

大地坐标（geodetic coordinate）　见“大地坐标系”．

大地纬度（geodetic latitude）　见“大地坐标系”．

大地经度（geodetic longitude）　见“大地坐标系”．

地心直角坐标系（geocentric rectangular coordinate system）　现代大地测量学采用的一种主要坐标系．即以地球质心为原点建立的空间直角坐标系．如图所示，地心直角坐标系 $Oxyz$ 的 z 轴与地球协议平自转轴重合，与 z 轴垂直的平赤道面构成 xy 平面，xz 平面是包含协议平自转轴和格林尼治平均天文台的平面；y 轴的指向使该坐标系成为右手坐标系．P 点的地心大地坐标 (B,L,H) 和该点地心直角坐标 (x,y,z) 之间存在着严格的数学关系，可以互相换算．



垂线偏差（deflection of the vertical）　描述地球重力场的一个几何量．即大地水准面上一点的垂线与该点椭球面法线间的夹角．它表示大地水准面相对参考椭球面的倾斜．垂线偏差通常用两个分量表示：一个是子午圈分量 ξ，即垂线偏差的南北分量；另一个是卯酉圈分量 η，即垂线偏差的东西分量．

垂线偏差可用于计算高程异常和大地水准面起伏、推算平均地球椭球或参考椭球的大小和形状及定位、归算天文大地测量观测数据，也可用于空间技术和精密工程测量．

因为平均地球椭球是惟一的，所以将垂线同平均地球椭球面法线间的夹角称为绝对垂线偏差，亦称重力垂线偏差．重力垂线偏差可应用维宁·曼尼茨公式计算．相对垂线偏差又称天文大地垂线偏差，是垂线和参考椭球面法线间的夹角．由于不同的参考椭球过地面点的法线不同，垂线偏差也不相同，故具有相对意义．相对垂线偏差可以利用天文经纬度和大地经纬度来计算．

绝对垂线偏差（absolute deflection of the vertical）　见“垂线偏差”．

重力垂线偏差（gravimetric deflection of the vertical）　见“垂线偏差”．

相对垂线偏差（relative deflection of the vertical）　见“垂线偏差”．

天文大地垂线偏差（astro-geodetic deflection of the vertical）　见“垂线偏差”．

平均地球椭球（mean earth ellipsoid）　亦称总地球椭球．地球形状理论中的一个基本概念．是一个与大地体最为接近的旋转椭球体．它满足三个条件：

1. 它的中心与地球的质心重合，它的短轴与地球平均自转轴重合，二者的旋转角速度相同；
2. 它的总质量等于地球的总质量；
3. 椭球面与大地水准面之间的差距平方和最小．

平均地球椭球是个数学形体，它的确定取决于地球重力场的确定和地心坐标系的建立，只有综合利用全球的地面大地测量、天文测量、重力测量和卫星大地测量资料，才能够精确求得与大地体（大地水准面包围的封闭体）最为接近的平均地球椭球．

弧度测量（arc measurement）　测定地球形状和大小的经典大地测量方法．即利用天文测量和大地测量结果推算地球形状、大小的方法．历史上，早期是测量地面上两段子午线弧长，并用天文测量方法测定诸端点的纬度，由此推算地球椭球的长半轴和扁率．中国清代康熙年间（公元 1708—1718 年）为编制《皇舆全图》，曾实施了大规模的天文大地测量．在这次测量中，证实高纬度的每度子午弧比低纬度的每度子午弧要长．此外，康熙还决定以每度子午弧长为 200 里（1 里＝0.5 千米）来确定里的长度．

大地水准面起伏（geoidal undulation）　亦称大地水准面高或大地水准面差距．是描述大地水准面形状的几何量．大地水准面与平均地球椭球面或参考椭球面之间（沿椭球面法线）的距离，都称为大地水准面起伏．前者是绝对的，也是惟一的；后者是相对的，随所采用的参考椭球及其定位的不同而异．

绝对大地水准面起伏的数值最大约为 ±100m．绝对大地水准面起伏可以利用全球重力异常，按斯托克斯公式进行数值积分算得；也可以利用地球重力场模型的位系数，按计算点的坐标求和算得；此外，还可以用卫星测高仪方法确定．

相对大地水准面起伏可以从大地原点出发，按天文水准或天文重力水准方法，计算出各点间的相对大地水准面起伏之差，然后逐段递推求得.

大地水准面高(geoidal height)　即"大地水准面起伏".

国家大地网(national geodetic network)　亦称国家大地控制网或天文大地网. 一个国家统一的大地点位测量控制网络. 是在全国领土范围内，由互相联系的大地点所构成的网. 大地点的水平位置按国家统一的作业方法测定，并设有固定标志，以便长期保存. 国家大地网一般采用三角测量方法，有时也兼用精密导线测量方法，采用由大到小、逐级控制的原则布设.

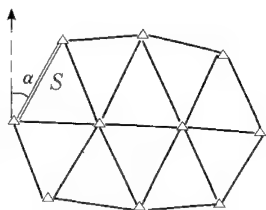
中国的国家大地网，除了青藏高原采用精密导线测量方法外，其他地区均用三角测量方法分成一、二、三、四等逐级布设. 一等三角测量是沿着经线和纬线方向布设成纵横交叉的三角锁系，以构成许多锁环. 1958 年以前，在一等三角测量锁环内布设二等三角锁和二等填充网. 1958 年以后，在一等三角测量锁环内布设二等全面网. 在二等网内再用三等和四等三角测量逐级加密. 为了控制角度观测误差积累的影响，在一等三角锁段的交叉处及锁环内布设若干起始边，测量其长度，并在端点上测定天文经度、纬度和方位角.

由于 GPS 定位技术的普及，各国都建立了国家 GPS 定位网，一般只分 A，B 两级. 其定位精度均高于用地面大地测量方法建立的国家大地网，GPS 定位网将可能逐步取代传统的大地控制网.

国家大地控制网(national geodetic control network)　即"国家大地网".

天文大地网(astro-geodetic network)　即"国家大地网".

三角测量(triangulation)　一种测量和推算地面点位的方法. 是在地面上按一定的技术要求布设一系列连续三角形，采取测角方式测定各三角形顶点(即三角点)水平位置的方法. 上述连续三角形可构成锁状或网状. 要测定网中所有的角度，并且至少还要测定一条边的长度 S 和一条边的方位角 α(见图)，从而可以由已知坐标的起始点出发推算出各点的坐标.



中国国家三角测量分为一、二、三、四等三角测量，各等级三角网的边长逐级缩短，观测精度也逐级降低(参见"国家大地网").

导线测量(traverse survey)　测量和推算地面点位的一种方法. 是在地面上按一定的技术要求选定一系列点(导线点)，连结相邻点成折线，依次测定各折线边(导线边)的长度和相邻边间的夹角(转折角或导线角)，根据起始点的坐标和起始方位角，利用测得的边长和转折角，逐点推算各导线点的坐标的方法.

为建立国家大地网，城市测量控制网和工程测量所实施的导线测量称为精密导线测量. 其等级和精度要求与相应等级的三角测量相同. 导线测量按距离测量的方法，可分为电磁波测距导线、视差导线、视距导线、经纬仪导线等. 导线测量的优点是：布设灵活，边长精度均匀. 缺点是控制面积小，缺乏有效和可靠的检核方法.

导线点(traverse point)　见"导线测量".

导线边(traverse leg)　见"导线测量".

国家水准网(national leveling network)　亦称国家高程控制网. 一个国家统一的地面点海拔高程测量控制网络. 在全国领土范围内，由水准测量测定高程的点(称水准点)所构成的网. 水准点是一种固定标志，以供长期保存，为国家经济建设、国防建设、地形测量和有关科学研究提供地面点高程.

中国国家水准网中，水准点的高程是由一、二、三、四等水准测量确定的，其施测精度逐级降低，由高级控制低级. 一等水准路线构成 100 个平均周长为 1500km 相互邻接的闭合环，在这些环中再用二、三、四等水准路线逐级加密，以青岛水准原点作为高程测量起算点，推算出所有水准点的高程.

国家高程控制网(national height control net)　即"国家水准网".

水准原点(leveling origin)　水准测量传递海拔高程的基准点. 即国家高程控制网中，所有水准点高程的起算点. 为了建立全国统一的高程控制网，必须确定一个高程起算面(水准基面)，作为所有水准点高程的起算基准. 通常采用大地水准面作为水准基面，它是沿海验潮站长期的海水面升降观测结果的平均值. 由于平均海面并不是重力等位面，不同地点的平均海面可相差 1～2m，中国采用设在青岛市的验潮站所确定的平均海面作为水准基面.

水准基面必须可靠地标定在地面上，为此在验潮站附近设置永久性水准原点，用精密水准测量测定原点对于验潮站平均海水面的高程. 中国的水准原点设在青岛市观象山上，它相对于 1952—1979 年期间青岛验潮站验潮观测资料所确定的黄海平均海面(即"1985 中国高程基准"的水准基面)的高程为 72.2604m.

高程系统(height system)　对地面点高程表示方法的定义. 采用不同的高程基准面表示地面点高低，称为定义了不同的高程系统. 大地测量中常遇到的高程系统一般有正高、大地高、正常高等高程系

统.

大地高(geodetic height) 高程系统的一种.即以参考椭球面作为高程基准面的一种高程系统.地面点的大地高为该点沿椭球面法线至参考椭球面的距离.三角高程测量观测结果经垂线偏差改正后,可以计算两点间的大地高高差.大地高又可直接由卫星大地测量方法测定.

正高(orthometric height) 高程系统的一种.即以大地水准面作为高程基准面的一种高程系统.地面点的正高为该点沿其垂线方向到大地水准面的距离.它的计算公式是:

$$H^g = \frac{1}{g_m}\int_{O'B} g\mathrm{d}h,$$

式中 $\mathrm{d}h$ 为沿水准路线测得的微分高差;g 为沿水准路线的重力值;g_m 是沿地面点 B 的垂线至大地水准面之间的平均重力值.由于 g_m 无法精确求定,通常采用近似的数据,故只能得到正高的近似值.

正常高(normal height) 高程系统的一种.即以似大地水准面作为高程基准面的一种高程系统.地面点的正常高为该点沿正常重力线到似大地水准面的距离,也等于似地球表面沿正常重力线到椭球面的距离.它的计算公式是

$$H^\gamma = \frac{1}{\gamma_m}\int_{OB'} g\mathrm{d}h,$$

式中 $\mathrm{d}h$ 为沿水准路线测得的微分高差;g 为沿水准路线的重力值;γ_m 是似地球表面点沿正常重力线到椭球体的正常重力平均值.中国各等级水准点的高程采用正常高高程系统.

似大地水准面(quasi-geoid) 一种接近大地水准面的几何面.是从地面点沿正常重力线向下截取该点的正常高,其端点所构成的曲面.似大地水准面很接近于大地水准面,在海洋上两者基本重合;在平原地区两者相差仅几厘米;在高山区两者最多相差4m.似大地水准面不是等位面,也没有明确的物理意义.

高程异常(height anomaly) 是描述似大地水准形状的几何量.即似大地水准面至平均地球椭球面的距离,也等于地球表面到似地球表面的距离.它是由扰动位产生的,一般用 ζ 表示.地面点的大地高 H 可以表示为正常高 H^γ 与高程异常 ζ 之和,即

$$H = H^\gamma + \zeta.$$

高程异常可以根据全球重力数据或采用天文水准、天文重力水准方法推算.

椭球面大地测量学(ellipsoidal geodesy) 大地测量学的一个分支.椭球面大地测量学将椭球作为地球的数学模型,研究椭球面上的测量计算、椭球面与其他曲面的关系及大地坐标系的建立等问题.

椭球面上的大地测量计算问题主要包括椭球面的数学性质、椭球面三角形解算、大地主题解算等.椭球面与其他曲面的关系包括:

1. 研究椭球面与地面的关系,以实现地面观测数据到椭球面的转化.

2. 研究椭球面与投影平面的关系、椭球面与大地水准面的关系、一个椭球面与另一个椭球面的关系,以实现不同大地测量坐标系统的变换.

子午圈(meridian) 椭球面上的一种平面截线.包含参考椭球短轴的平面称为子午面,子午面与参考椭球面的截线称为子午圈.

子午面(meridian plane) 见"子午圈".

子午圈曲率半径(radius of curvature in meridian) 描述椭球面曲率的几何量.即地球椭球面上某点的主曲率半径之一,它是该点所有法截线曲率半径中最小的曲率半径.在大地坐标系中,其计算公式为

$$M = \frac{a(1-e^2)}{W^3}$$

或

$$M = \frac{N}{V^2},\quad M = \frac{C}{V^3},$$

式中 M 为子午圈曲率半径,a 为椭球长半径,N 为卯酉圈曲率半径,

$$W = (1 - e^2\sin^2 B)^{1/2},$$
$$V = (1 + e'^2\cos^2 B)^{1/2},$$

e 和 e' 分别为椭球第一和第二偏心率,B 为点的大地纬度.M 随 B 增大而增大,在赤道上,

$$M_0 = a(1-e^2) < a;$$

在极点上,

$$M_{90} = \frac{a}{\sqrt{1-e^2}} = C,$$

C 是极点曲率半径.子午圈曲率半径在子午线弧长计算、平均曲率半径计算以及理论推导中有广泛应用.

卯酉圈(prime vertical) 椭球面上的一种平面截线.参考椭球面上包含一点的法线且与该点子午面正交的平面称为卯酉面,它与椭球面的截线称为卯酉圈.

卯酉面(prime vertical plane) 见"卯酉圈".

卯酉圈曲率半径(radius of curvature in prime vertical) 描述椭球面曲率的几何量.即椭球面上某点的主曲率半径之一,是该点所有法截线曲率半径中最大的曲率半径.在大地坐标系中,其计算公式为

$$N = \frac{a}{W} \text{ 或 } N = \frac{C}{V},$$

式中 a 为椭球长半径,C 为极点曲率半径(参见"子午圈曲率半径"),

$$W = (1 - e^2\sin B)^{1/2},$$

$$V=(1+e'^2\cos^2 B)^{1/2},$$

e 和 e' 分别为椭球第一和第二偏心率，B 为点的大地纬度. N 随 B 增大而增大，在赤道上，$N_0=a$；在极点上，$N_{90}=C$. N 值恰好等于法线在椭球面与短轴间的长度. 卯酉圈曲率半径在各种椭球面计算和理论推算中具有广泛的应用. 由于 N 的几何意义明显，在适当的范围内常把地球椭球面当做以它为半径的球面，以简化计算.

平均曲率半径(mean radius of curvature)　描述椭球面曲率的几何量. 即椭球面上一点所有方向法截线曲率半径的平均值. 旋转曲面上点的平均曲率半径等于该点两主曲率半径乘积的平方根(即几何中数)，地球椭球面是一个旋转曲面，该曲面上的一点的平均曲率半径

$$R=\sqrt{MN},$$

其中 M 和 N 分别为子午圈和卯酉圈曲率半径. 在测量与制图中，常用某点的平均曲率半径 R 为半径的球面来代替该点附近的地球椭球面.

法截线(normal section curve)　椭球面上一种平面截线. 即包含椭球面上一点法线的平面(法截面，又称法平面)与椭球面的截线. 过椭球上一点可作无数法截线，方位角为 0°和 180°的(主)法截线就是子午圈；方位角为 90°和 270°的(主)法截线就是卯酉圈.

法截线在大地测量中有重要作用. 地面上观测方向是以垂线为基准的，经垂线偏差改正后的照准面与椭球面的截线就是法截线.

大地线(geodesic)　曲面上的短程线. 在大地测量中，指地球椭球面上两点间最短的连线. 大地线上每点的密切平面都包含该点的曲面法线. 地球椭球面上的法截线(参见“法截线”)通常不是大地线，只有子午线和赤道是大地线. 在大地测量计算中，任意两点间的边长和方向都以大地线为准.

大地线微分方程(differential equation of geodesic)　一种短程线微分方程. 即地球椭球面上大地线长度与大地经度、纬度、大地方位角间的微分关系式. 设大地线长度为 S，大地经度、纬度分别为 L 和 B，大地方位角为 A，则大地线微分方程如下：

$$\mathrm{d}B=\frac{\cos A}{M}\mathrm{d}S,$$

$$\mathrm{d}L=\frac{\sin A}{N}\sec B\mathrm{d}S,$$

$$\mathrm{d}A=\frac{\sin A}{N}\tan B\mathrm{d}S,$$

式中 M,N 分别为子午圈和卯酉圈曲率半径. 大地线微分方程在理论推导中具有重要意义.

大地方位角(geodetic azimuth)　亦称拉普拉斯方位角. 一种在大地坐标系中表示方向的角量. 是参考椭球面上过某点的子午圈与过该点某一方向的大地线间的夹角. 大地方位角由子午圈北方向起按顺时针方向计算，通常用 A 表示，它不能直接测得，而是由天文方位角按拉普拉斯方程换算而得. 拉普拉斯方程为

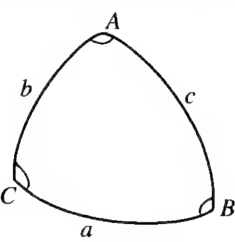
$$A=\alpha-(\lambda-L)\sin\varphi,$$

式中 α,λ,φ 分别是天文方位角、天文经度、天文纬度；L 为大地经度.

拉普拉斯方位角(Laplace azimuth)　即“大地方位角”.

拉普拉斯点(Laplace point)　一种国家大地网点. 即在国家大地网中测有天文经度、天文纬度和天文方位角的点称为拉普拉斯点.

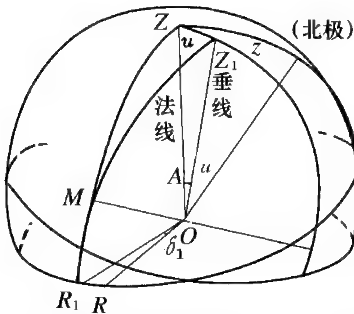
勒让德定理(Legendre theorem)　解算球面三角形边长的一种方法. 若平面三角形和球面三角形对应边长相等，则平面三角形的角等于对应球面三角形的角减去球面角超的三分之一. 勒让德定理的特点是不改变球面三角形边长，而是改变它的角度，运用平面三角公式解算，得到的边长就是球面三角形边长. 球面角超 ε 的计算公式如下：



$$\varepsilon=\frac{bc\sin A}{2R^2},$$

式中 R 为球面曲率半径，其他符号意义见之于图.

垂线偏差改正(correction for deflection of the vertical)　地面大地测量的一种改正. 即垂线偏差对水平方向观测影响的改正. 地面上以垂线为准观测的水平方向归算为参考椭球面上以法线为准的水平方向值时，顾及垂线偏差影响所加的改正数，通常用 δ_1 表示之. 设 ξ,η 分别为垂线偏差 u 在子午圈和卯酉圈上的分量，图中，OM 为测站 O 对地面目标的观测方向，OR_1 和 OR 分别为 OM 以垂线为准和以椭球法线为准的水平方向，A 为观测方向的大地方位角，z 为观测目标的天顶距，则有



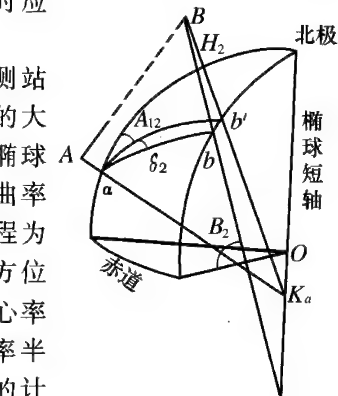
$$\delta_1=-(\xi\sin A-\eta\cos A)\cot z,$$

δ_1 的值一般小于 0.1″，但在一、二等三角测量中必

须顾及此项改正.

标高差改正(correction for skew normal)　地面大地测量的一种改正. 水平方向观测中,因照准点不在椭球面上,归算时应施加的一种改正.

在图中,设 A 为测站点,B 为照准点,B 点的大地纬度为 B_2,该点在椭球面投影点 b 的子午圈曲率半径为 M_2,照准点高程为 H_2,观测方向的大地方位角为 A_{12},椭球第一偏心率为 e, M 为子午圈曲率半径,则标高差改正 δ_2 的计算公式为



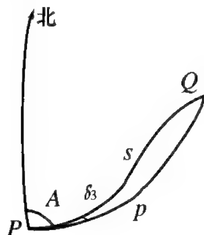
$$\delta_2 = H_2(1)_2 \frac{e^2}{2}\sin 2A_{12} \cos^2 B_2,$$

其中 $(1)_2 = \rho''/M_2$. δ_2 一般不大于 0.1″. 在一、二等三角测量中必须顾及标高差改正.

截面差改正(correction from normal section to geodesic)　地面大地测量的一种改正. 即将法截线方向归算为大地线方向应加的改正. 在图中设 P 为测站,Q 是照准点,$\overset{\frown}{PpQ}$ 为 P 至 Q 的法截线,$\overset{\frown}{PsQ}$ 是 P,Q 间的大地线,它们间的夹角 δ_3 就是截面差改正.

$$\delta_3 = -\frac{S^2\rho''}{12N_1^2}\eta_1^2 \sin 2A,$$

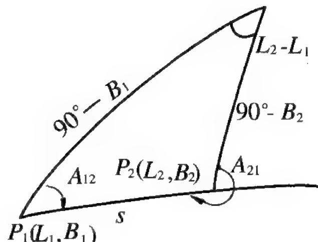
式中 S 为 $\overset{\frown}{PsQ}$ 的长度,N_1 为按 P 点的大地纬度算得的卯酉圈曲率半径,A 为观测方向的大地方位角.



$$\eta_1^2 = e'^2\cos^2 B_1,$$

其中 e' 为椭球的第二偏心率,B_1 为 P 点的大地纬度. 截面差改正数很小,只在一等三角测量中才需要顾及它.

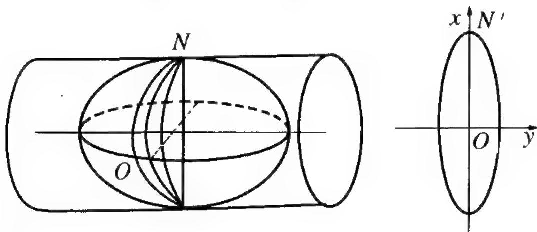
大地主题解算(solution of geodetic problem)　关于在参考椭球面上推算大地坐标、大地线长度和大地方位角的解算方法. 依推算元素的不同,可分为大地主题正解和大地主题反解. 如图,已知 P_1 点的大地坐标为 $P_1(L_1,B_1)$,P_1 点至 P_2 点的大地线长度为 s,大地方位角为 A_{12},推算 P_2 点的大地坐标 $P_2(L_2,B_2)$ 和反大地方位角 A_{21},称大地主题正解;已知 P_1, P_2 两点的大地坐标分别为 $P_1(L_1,B_1)$ 和 $P_2(L_2,B_2)$,反算大地线长度 s 和大地方位角 A_{12},A_{21},称为大地主题反解. 由图可知,大地主题解算实质上是一个解算椭球面上的极三角形问题. 具有代表性的解算方法有:勒让德级数、高斯中纬度公式和贝塞尔公式等. 若按其解算距离分类,大致可分为短距离($s <$ 400 km)、中距离(400km $\leqslant s <$ 1000 km)和长距离($s >$ 1000 km)的大地主题解算. 大地主题解算主要用于一等三角测量计算、洲际联测、空间飞行器发射和无线电导航等.



高斯-克吕格投影(Gauss-Krüger projection)　简称高斯投影. 是地球椭球面与平面间正形投影的一种方法. 它由高斯(Gauss,C. F.)于 1820 年创立,后经克吕格(Krüger,I. H. L.)对投影公式加以补充和完善. 该投影从地球椭球面正形投影于平面的一般公式

$$x + \mathrm{i}y = f(q + \mathrm{i}l) \quad (\mathrm{i}^2 = -1)$$

出发,按照投影带中央的子午线(中央子午线)的描写形为直线,且长度不变和赤道投影后为直线的条件,确定投影函数的表达式,从而得到高斯-克吕格投影公式. 由此可以将大地坐标转换为平面直角坐标(高斯-克吕格坐标). 如图,投影后,除中央子午线和赤道为直线外,其他子午线均为对称于中央子午线的曲线. 这一投影方法相当于用一个椭圆柱横切于椭球面上投影带的中央子午线上,按上述条件将中央子午线两侧一定经度差范围内的椭球面正形投影于椭圆柱面,展开椭圆柱面即得高斯投影平面. 取中央子午线与赤道交点的投影为原点,中央子午线的投影为纵坐标轴 x;赤道的投影为横坐标轴 y,于是构成高斯-克吕格平面直角坐标系. 此投影离中央子午线愈远的区域,投影后长度变形愈大,故必须采用分带(参见“高斯投影分带”)的方式加以限制. 高斯投影具有投影公式简单、各带投影公式相同等优点,已为中国和其他许多国家采用.



高斯投影分带(zone dividing of Gauss projection)　高斯投影中限制长度变形的基本方法. 按一定经度差将地球椭球面划分成若干瓜瓣形投影带,以便分带投影. 通常根据测图比例尺的要求,按 6°或 3°经度差分为六度带或三度带. 六度带自 0°子午线起每隔 6°经度差自西向东分带,带号依次编为

1,2,…,60. 6°中央子午线经度 $L = 6n - 3$,其中 n 为带号.三度带是在六度带的基础上划分的,它的中央子午线与六度带的中央子午线和分带子午线重合,即自 1.5° 子午线起,每隔 3° 经度差自西向东分带,带号依次编为 1,2,…,120.三度带中央子午线经度 $L = 3n$,式中 n 为三度带带号.中国的经度范围西起 73° 东至 135°,分成六度带 11 带或三度带 21 带.六度带用于中小比例尺测图,三度带可用于大比例尺测图.

高斯平面坐标(Gauss plane coordinate) 地图投影中的一种平面直角坐标.高斯-克吕格投影是分带投影的,故各带坐标成独立系统.各带以中央子午线的投影为纵轴(x),赤道投影为横轴(y),两轴的交点即为各带的坐标原点.纵坐标以赤道为零起算,赤道以北为正,以南为负;横坐标若从中央子午线起算,将出现负值,使用不便.为此,规定将纵坐标轴西移 500km,即凡是带内的横坐标值均加 500km.为了标明某一坐标系统所属投影带,在加了 500km 的横坐标值前再冠以带号.

高斯平面子午线收敛角(Gauss grid convergence) 描述子午线投影向南北两极收敛弯曲的角量.如图,即过投影平面上某点 P 平行于纵坐标轴的直线与其子午线投影曲线间的夹角.在高斯投影平面上,这个夹角称为高斯平面子午线收敛角,通常以 γ 表示.在高斯投影中,除中央子午线投影为直线外,其两侧的子午线投影后均为弯向纵坐标轴的曲线,故除中央子午线外,其他各点处均有 γ 存在.若 P 点在中央子午线以东,则 γ 为正,否则为负.子午线收敛角主要用于大地方位角与坐标方位角的换算.

高斯投影方向改正(arc-to-chord correction in Gauss projection) 亦称高斯投影曲率改正.简称方向改正.一种地图投影改正.即地球椭球面上两点间的大地线方向换算为高斯投影平面上相应两点间的直线方向所加的改正,分别用 δ_{12} 和 δ_{21} 表示.将 δ 加于大地线方向值上,即得高斯投影平面上直线的方向值.精度为±0.1″的计算公式为

$$\delta_{12} = -\delta_{21} = \frac{\rho''}{2R_m^2}(x_2 - x_1)y_m,$$

式中

$$\rho'' = \frac{180}{\pi} \times 3600'',$$

x_1,x_2 分别为高斯投影平面上线段两端点的纵坐标,y_m 为两端点横坐标的平均值,R_m 为该线中点处地球椭球体的平均曲率半径.

高斯投影曲率改正(curvature correction in Gauss projection) 即"高斯投影方向改正".

方向改化(arc to chord correction) 高斯投影曲率改正的简称.

高斯投影距离改正(distance correction in Gauss projection) 一种地图投影改正.高斯投影中投影距离的归算改正.即将地球椭球面上两点间的大地线长度,换算为高斯投影平面上相应两点间的直线距离时所加的改正.设大地线长度为 S,它在高斯投影平面上的长度为 s,相应的直线距离为 D.当距离 D 小于 70km 时,由于 s 与 D 之差小于 1mm,故在实际计算中只要求出 $\Delta S = s - S$,将 ΔS 加于 S 中即得投影平面上的直线距离 D.计算精度达 ±1mm 的计算公式为

$$\Delta S = S\left(\frac{y_m^2}{2R_m^2} + \frac{\Delta y^2}{24R_m^2} - \frac{y_m^4}{24R_m^4}\right),$$

式中 y_m 为高斯平面上直线段两端点横坐标的平均值,Δy 为其横坐标之差,R_m 为直线中点处的地球平均曲率半径.

物理大地测量学(physical geodesy) 亦称大地重力学.大地测量学的主要分支之一.是一门研究地球形状和地球外部重力场及其在大地测量中应用的学科.主要任务是:利用地面重力测量和卫星重力探测方法建立全球或区域性地球重力场模型并确定大地水准面,据此确定一个与地球真实形状或大地水准面最接近的旋转椭球,即平均地球椭球,作为建立大地坐标系的参考面,并为大地测量观测量的归算提供重力场参数.主要研究内容包括:地面重力测量技术、卫星重力探测理论和技术、重力归算理论、重力位理论、重力场边值问题、正常重力场和正常椭球理论、重力异常场或扰动场理论、确定大地水准面和垂线偏差的方法等.

20 世纪 70 年代发展起来的卫星重力探测技术,由于其全球性高重复探测能力,提供了丰富的地球重力场信息,大大改善了全球重力场模型的分辨率和准确度,为相关地学学科的研究和航天技术的发展提供日益精细化的重力场结构.这些相关学科也在推动物理大地测量学的发展,特别是卫星重力技术,正在成为研究地球形状和地球外部重力场的主要方法和技术.

大地重力学(geodetic gravimetry) 即"物理大地测量学".

重力(gravity) 物理大地测量学研究的对象.即地球上单位质点所受万有引力与离心力之和.采用加速度的单位,通常为 cm/s²,称为伽(gal),10^{-3}伽称为毫伽(mgal),10^{-3}毫伽称为微伽(μgal)(加速度的国际法定单位为 m/s²,1gal $= 10^{-2}$m/s²,国际计量局将单位伽(包括毫伽、微伽)列为暂时与

国际单位制并用的单位).地球上单位质点所受的万有引力包括地球以及所有其他天体的引力.但在研究地球重力场时,均假设这个引力只是地球的引力.日、月的引力和地球绕地月中心旋转而产生的惯性离心力之和称为重力固体潮,其值很小,其余天体对地球的引力可忽略不计.在物理大地测量中,一般是在重力观测值中去掉重力固体潮,只研究与重力固体潮无关的部分.

重力固体潮(gravity earth tides) 见“重力”.

重力测量(gravity measurement) 物理大地测量的主要测量方法.即测定重力值的技术和方法.它分为绝对重力测量和相对重力测量.绝对重力测量是测定重力的绝对量值.目前使用的方法是利用激光干涉仪和原子钟分别测定自由落体或抛物体的运动距离和时间来计算重力,精度可达几个微伽($10^{-8}m/s^2$).测定绝对重力的仪器称绝对重力仪.相对重力测量是测定不同位置间的重力之差,目前使用的仪器均为弹簧结构,称相对重力仪或重力仪,精度可达10微伽($10^{-7}m/s^2$).重力仪一格读数对应的重力值称格值.使用前必须对每台重力仪的格值进行测定,称之为重力仪的标定.重力仪在同一点的读数会随时间变化,这种现象称为零点漂移.为了消除零点漂移,必须进行往返测量或对已知点-未知点-已知点联测.

有的重力仪是不可移动的,这种仪器主要用来观测某一点处重力随时间的变化,例如超导重力仪.

绝对重力测量(absolute gravimetry) 见“重力测量”.

相对重力测量(relative gravimetry) 见“重力测量”.

重力点(gravimetric point) 已测重力值的点.

绝对重力点(absolute gravity point) 重力点的一种.即国家重力基准网中进行过绝对重力测量的点称为绝对重力点.

重力位(gravity potential) 重力场理论中的一个基本概念.即描述重力场势能的物理量,是一个标量函数,定义其梯度等于重力.它的量纲与单位质点位能的量纲一样,为m^2/s^2或cm^2/s^2.重力位为引力位与离心力位之和,引力位V的梯度为万有引力,离心力位Q的梯度为离心力.V和Q的数学表达式分别为

$$V = G\int \frac{\mathrm{d}m}{r}, \quad Q = \frac{1}{2}d^2\omega^2,$$

式中G为万有引力常数,r为计算点到吸引质元$\mathrm{d}m$的距离,积分是对吸引质量所占空间进行的,可以是线、面或体等,d计算点为到自转轴的距离,ω为自转角速率.质体引力位在质体外部满足拉普拉斯方程$\Delta V=0$,在质体内部满足泊松方程

$$\Delta V = -4\pi G\rho,$$

式中ρ为密度.离心力满足泊松方程$\Delta Q = 2\omega^2$.

地球重力场(gravity field of the earth) 物理大地测量学的主要研究对象.即地球万有引力和离心力合力的力场.研究地球重力场主要研究地球重力位,它包含了地球重力场中所具有的所有信息,其中扰动位、重力异常、大地水准面差距、垂线偏差等是地球重力场的基本参数.确定地球重力场就是确定这四个基本参数,并研究它们的应用.一般是将引力位展开成球函数级数来表示地球重力场,称为地球重力场模型,

$$V(r,\theta,\lambda) = \frac{GM}{r}\left\{1 + \sum_{n=2}^{N}\left(\frac{a}{r}\right)^n (A_n^m\cos m\lambda + B_n^m\sin m\lambda)\mathrm{P}_n^m(\cos\theta)\right\},$$

式中GM为万有引力常数与地球质量的乘积,(r,θ,λ)为以地球质心为中心、地球转动惯量最大的(即通过极点的)主惯性轴为极轴的球坐标,a为平均椭球体的赤道半径.A_n^m和B_n^m称为位系数,$\mathrm{P}_n^m(\cos\theta)$为勒让德函数,或称勒让德缔合函数.

地球重力场模型(earth gravity field model) 见“地球重力场”.

正常重力位(normal gravity potential) 重力场理论中的一个基本概念.即一种正常地球模型的重力位,或平均地球椭球的重力位.其球函数级数表示式为

$$U(r,\theta) = \frac{GM}{r}\left[1 - \sum_{n=1}^{\infty}J_{2n}\left(\frac{a}{r}\right)^{2n}\mathrm{P}_{2n}(\cos\theta) + \frac{1}{2}\omega^2r^2\sin^2\theta\right],$$

式中系数

$$J_{2n} = (-1)^{n+1}\frac{3e^{2n}}{(2n+1)(2n+3)} \cdot \left(1 - n - 5n\frac{C - \frac{A+B}{2}}{Ma^2e^2}\right),$$

G,M和ω分别为万有引力常数、地球的质量、地球自转角速率;A,B,C分别为主惯量,其中C是绕地心坐标极轴(z轴)的转动惯量;a和e分别是平均地球椭球体的赤道半径和第一偏心率;(r,θ)为子午平面地心球坐标,由于旋转椭球的对称性,U与经度λ无关.

正常重力(normal gravity) 描述地球重力场规则部分的一种模型重力.即平均地球椭球,或称正常椭球的重力,是正常重力位的梯度.若用γ_e表示赤道上的重力,γ_p表示极点的重力,a,b分别表示平均地球椭球的长、短半轴,则平均地球椭球面上地理纬度B处的正常重力为

$$\gamma_0 = \frac{a\gamma_e\cos^2B + b\gamma_p\sin^2B}{\sqrt{a^2\cos^2B + b^2\sin^2B}}.$$

这一公式是索米利亚纳(Somigliana,C.)于 1929 年导出的,称为索米利亚纳正常重力公式.在应用中,一般将上式展开成级数:

$$\gamma_0 = \gamma_e(1 + \beta\sin^2 B - \beta_1\sin^2 2B),$$

式中

$$\beta = \frac{\gamma_p - \gamma_e}{\gamma_e},$$

称为重力扁率,

$$\beta_1 = \frac{1}{2\alpha^2} + \frac{1}{4\alpha\beta}$$

(α 为平均椭球体的扁率).在地球的正常重力场模型中,一般给出 γ_e,β 和 β_1 三个值.

索米利亚纳正常重力公式(Somigliana normal gravity formula) 见"正常重力".

重力扁率(gravity flattening) 见"正常重力".

地球动力因子(dynamic factor of the earth) 地球形状学的一个基本物理量.即正常重力位中的 J_2:

$$J_2 = \frac{C - \dfrac{A+B}{2}}{Ma^2},$$

式中各量的物理意义参见"正常重力位".在确定平均椭球体时,选择 $J_2 = -A_2^0$,这里 A_2^0 是地球重力场的位系数.

克莱罗公式(Clairaut formula) 地球形状学的基本公式.即由重力测量计算平均椭球体形状的公式,即

$$\alpha + \beta = \frac{\omega^2 b}{\gamma_e}\left\{1 + \frac{e'\left[3\left(1+\dfrac{1}{e'^2}\right)(1 - e'\arctan e') - 1\right]}{\left(1+\dfrac{3}{e'^2}\right)\arctan e' - \dfrac{3}{e'}}\right\},$$

式中 α,β,b,γ_e 的意义见"正常重力",ω 和 e' 分别为地球的自转角速率和平均椭球体的第二偏心率.这一个公式是由克莱罗(Clairaut,A.-C.)于 1738 年提出的.它的级数形式为

$$\alpha + \beta = \frac{5}{2}q - \frac{26}{7}\alpha q + \frac{15}{4}q^2,$$

式中 $q = \omega^2 a^3/GM$,G,M 和 a 的意义见"正常重力位".

由于现代卫星技术可直接精确测定地球动力因子 J_2,因此将克莱罗公式改写成

$$J_2 = \frac{1}{3}e^2\left\{1 - \frac{4qe'(1-\alpha)}{15\left[\left(1+\dfrac{3}{e'^2}\right)\arctan e' - \dfrac{3}{e'}\right]}\right\},$$

式中 e 为平均椭球体的第一偏心率.该式的级数展开式为

$$\alpha = \frac{1}{2}q + \frac{3}{2}J_2 + \frac{9}{8}J_2^2 - \frac{3}{14}J_2 q - \frac{11}{56}q^2,$$

这是计算 α 的实用公式.

重力异常(gravity anomaly) 地球重力场的一个基本概念.它描述由于地球内部密度相对正常密度的异常所产生的重力异常.即大地水准面上某点的重力与该点在平均椭球面上投影处的正常重力之差

$$\Delta g = g_0 - r_0.$$

空间某一点上的重力与同一点正常重力之差称为重力扰动或纯重力异常.

实际上,重力是在地面上观测的,将地面上的重力值归算到大地水准面上时,可有不同的归算方法,因而可得到物理意义不同的重力异常(参见"重力归算").

在应用中,重力异常是按格网平均值给出的,例如,根据重力测量的精度和重力点的密度可给出 10°×10°,5°×5°,1°×1°等的全球或区域格网平均重力异常.

重力扰动(gravity disturbance) 见"重力异常".

纯重力异常(pure gravity anomaly) 见"重力异常".

扰动位(disturbing potential) 地球重力位理论中的基本概念.即地球上一点实际重力位与该点正常重力位之差.由于正常重力位可精确求出,并且非常接近重力位,所以求解重力位的问题可简化成求解扰动位的问题,而且扰动位是一微小量,并满足拉普拉斯方程.扰动位是解算以重力测量基本微分方程为边界条件、大地水准面为边界的拉普拉斯方程得到的.其具体计算见"布隆斯公式"和"斯托克斯公式".

重力测量基本微分方程(gravimetry basic differential equation) 物理大地测量学中的基本方程.即大地水准面上的扰动位与重力异常之间的关系式,

$$\Delta g = -\frac{\partial T}{\partial r} - \frac{2}{R}T,$$

式中 Δg 为重力异常,T 为扰动位,$\frac{\partial}{\partial r}$为对地心距离的偏导数,$R$ 为大地水准面的平均半径.这是由拉普拉斯方程解算扰动位 T 所需的边界条件.

布隆斯公式(Bruns formula) 物理大地测量学中的基本公式.即由扰动位计算大地水准面起伏的公式,即

$$N = \frac{T_0}{\gamma},$$

式中 γ 为平均椭球体表面的正常重力,可用全球平均值 $\bar{\gamma}$ 代替;T_0 为大地水准面上的扰动位.

斯托克斯公式(Stokes formula) 物理大地测

量学中的基本公式. 即由重力异常 Δg 计算大地水准面起伏 N 的公式,即

$$N=\frac{R}{4\pi\bar{\gamma}}\int_{\sigma}\Delta g S(\psi)\mathrm{d}\sigma,$$

其中 $\bar{\gamma}$ 为全球平均正常重力;R 为地球的平均半径;σ 为单位球面;ψ 为计算 N 的点和积分流动点之间的地心角;$S(\psi)$ 为斯托克斯函数,

$$S(\psi)=\frac{1}{\sin\frac{\psi}{2}}-6\sin\frac{\psi}{2}+1-5\cos\psi-3\cos\psi\ln\left(\sin\frac{\psi}{2}+\sin^2\frac{\psi}{2}\right).$$

上面公式是斯托克斯(Stokes,G. G.)于 1849 年导出的,称为斯托克斯公式. 斯托克斯公式与斯托克斯级数是等价的. 斯托克斯级数是

$$N=\frac{R}{\gamma}\sum_{n=2}^{\infty}\sum_{m=0}^{n}(A_n^m\cos m\lambda+B_n^m\sin m\lambda)\mathrm{P}_n^m(\cos\theta),$$

系数为

$$A_n^0=\frac{2n+1}{4\pi(n-1)}\int_{\sigma}\Delta g\mathrm{P}_n(\cos\theta')\mathrm{d}\sigma',$$

$$A_n^m=\frac{2n+1}{2\pi(n-1)}\int_{\sigma}\Delta g\mathrm{P}_n^m(\cos\theta')\cos m\lambda'\mathrm{d}\sigma'\ (m>0),$$

$$B_n^m=\frac{2n+1}{2\pi(n-1)}\int_{\sigma}\Delta g\mathrm{P}_n^m(\cos\theta')\sin m\lambda'\mathrm{d}\sigma'\ (m>0).$$

斯托克斯公式和斯托克斯级数是首先解算以重力测量基本微分方程为边界条件、以大地水准面(解算时近似地当做球面)为边界的拉普拉斯方程得到扰动位 T,然后由布隆斯公式得到的.

维宁·曼尼茨公式(Vening Meinesz formula) 物理大地测量学中的基本公式. 即由重力异常 Δg 计算重力垂线偏差的公式,即

$$\begin{cases}\xi''=-\dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}\mathrm{d}A\int_0^{\pi}\Delta g Q(\psi)\cos A\mathrm{d}\psi,\\ \eta''=-\dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}\mathrm{d}A\int_0^{\pi}\Delta g Q(\psi)\sin A\mathrm{d}\psi,\end{cases}$$

ξ''和 η''分别为子午圈和卯酉圈方向的垂线偏差分量,A 是垂线偏差的计算点到积分流动点连线的方位角,ψ 为这两点对应的地心角.

$$Q(\psi)=\frac{\rho''}{2\bar{\gamma}}\cos^2\frac{\psi}{2}\left[\csc\frac{\psi}{2}+12\sin\frac{\psi}{2}-32\sin^2\frac{\psi}{2}+\frac{3}{1+\sin\frac{\psi}{2}}-12\sin^2\frac{\psi}{2}\ln\left(\sin\frac{\psi}{2}+\sin^2\frac{\psi}{2}\right)\right],$$

式中 ρ''为每弧度包含的角秒数,$\bar{\gamma}$ 为全球平均正常重力. 上述公式是由维宁·曼尼茨(Vening Meinesz,F. A.)于 1928 年导出的.

重力归算(gravity reduction) 重力测量数据处理中的概念. 即重力场计算中对地面重力观测值所应作的归算. 在利用斯托克斯公式计算大地水准面差距或利用维宁·曼尼茨公式计算垂线偏差时,要求将地面的重力观测值归算到大地水准面上,并使大地水准面外没有质量,该过程称为重力归算.

空间异常(free-air anomaly) 重力异常场的一个基本概念. 将大地水准面外部的质量向下平移到大地水准面下,然后按地球外部的重力梯度将地面上的重力归算到大地水准面上,由此得到的重力异常称为空间异常,又称自由空气异常,其计算公式为

$$\Delta g=g+\Delta g_1-\gamma_0,$$

式中 g 为地面实测重力,γ_0 为平均椭球体上的正常重力,

$$\Delta g_1=0.3086H,$$

称为空间改正,又称自由空气改正,以毫伽(1mgal $=10^{-5}\mathrm{m/s^2}$)为单位,H 为海拔高程,以米为单位.

空间改正(free-air correction) 见"空间异常".

布格异常(Bouguer anomaly) 重力异常场的一个基本概念. 将大地水准面外部的质量去掉,然后将原来地面上的重力归算到大地水准面上得到的重力异常称为布格异常. 若不考虑地形而假设地面与大地水准面平行,则与此对应的布格异常称为不完全的布格异常,其计算公式为

$$\Delta g=g+\Delta g_1+\Delta g_2-\gamma_0,$$

式中 $\Delta g_2=-0.1118H$ 为层间改正,其余各量的意义与"空间异常"中的一样.

不完全布格异常(partial Bouguer anomaly) 见"布格异常".

地形改正(terrain correction) 一种重力归算改正. 即在布格异常的计算中考虑局部地形起伏的重力改正. 按下式计算的重力异常称为完全布格异常:

$$\Delta g=g+\Delta g_1+\Delta g_2+\Delta g_3-\gamma_0,$$

Δg_3 称为地形改正,是一个复杂的积分. 若只加空间改正和地形改正,则得到的重力异常称为法伊异常(式中各量意义见"布格异常"和"空间异常").

法伊异常(Faye anomaly) 见"地形改正".

均衡改正(isostatic correction) 一种重力归算改正. 即在重力归算中考虑地壳质量均衡效应的重力改正. 任意地面高出的部分都由地下的质量亏损所补偿,将大地水准面外部的质量填补到大地水准面内部质量亏损的区域后得到一项重力改正 Δg_4,称为均衡改正. 将均衡改正加到布格异常上得到的重力异常称为均衡异常. 均衡改正也是一个复杂的积分,因均衡模型的不同而不同.

均衡异常(isostatic anomaly) 见"均衡改正".

似地球表面(telluroid) 亦称地形表面. 确定地球真实形状的莫洛金斯基理论中的一个概念. 一

种根据正常高和天文经度、天文纬度定义的接近地球真实表面形状的曲面. 由平均椭球面起,沿法线方向量取正常高得到一点,由天文经度和天文纬度确定所有这样的点并连在一起构成的曲面称为似地球表面.

地形表面(telluroid) 即“似地球表面”.

地面重力异常(terrestrial gravity anomaly) 确定地球真实形状的莫洛金斯基理论中的一个概念. 地球表面上一点的重力异常(Δg),即地面实测重力和似地球表面上相应点的正常重力之差,其数值与大地水准面上的空间异常近似相等. 它与扰动位 T 的关系为

$$\Delta g = -\frac{2T}{r} - \frac{\partial T}{\partial r},$$

式中 r 为地面到地心的距离.

莫洛金斯基级数(Molodensky series) 莫洛金斯基理论中的基本公式. 即由地面重力异常计算地面扰动位的公式,即

$$T_P = T_0 + T_1 + \cdots,$$

式中的各项分别为

$$T_0 = \frac{R}{4\pi}\int_\sigma \Delta g S(\psi)\mathrm{d}\sigma,$$

$$T_1 = \frac{R}{4\pi}\int_\sigma\left[\frac{R^2}{2\pi}\int_\sigma \Delta g\,\frac{H^\gamma - H_P^\gamma}{r^3}\mathrm{d}\sigma\right]S(\psi)\mathrm{d}\sigma,$$

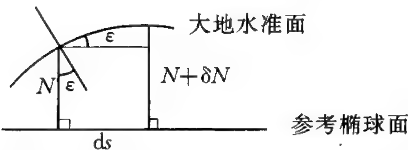
式中 H^γ 和 H_P^γ 分别为流动点和计算点 P 的正常高,其他符号的意义与“斯托克斯公式”中一致.

上述计算地面扰动位的级数公式是由莫洛金斯基(Молоденский, M. C.)于 1962 年导出的. 另外还有许多其他计算地面扰动位的公式,都是解算以地面重力异常和扰动位的微分关系为边界条件、以地面为边界的拉普拉斯方程得到的. 利用地面上的扰动位可求得高程异常和地面上的垂线偏差,从而求得地面相对于平均椭球体表面的位置,即地球表面真实形状.

天文重力水准(astro-gravimetric leveling) 一种利用天文和大地观测并结合重力测量推算大地水准面起伏的方法. 利用天文和大地观测可算出大地水准面上的天文大地垂线偏差,由此可求得相邻两点相对于参考椭球面的大地水准面起伏之差. 由图可见,

$$\delta N = \varepsilon \cdot \mathrm{d}s,$$

这种确定大地水准面的方法称为天文水准.



若已知天文大地垂线偏差的点相距较远,则可根据重力资料内插天文大地垂线偏差,然后利用天文水准的方法计算大地水准面,这种确定大地水准面的方法称为天文重力水准.

天文水准(astronomical leveling) 见“天文重力水准”.

卫星重力学(satellite gravimetry) 利用卫星技术研究地球重力场的一门学科. 主要是利用卫星轨道资料、卫星跟踪卫星资料、卫星测高资料以及卫星重力梯度资料等来研究地球的重力场. 目前,万有引力常数与地球质量的乘积 GM 以及地球动力学因子 J_2 都是利用卫星轨道求得的.

地心引力常数(geocentric gravitational constant) 描述地心引力场的特征参数. 通常用 K 表示,K 等于牛顿引力常数 G 乘以地球质量 M,即 $K=GM$. 若按 20 世纪 80 年代国际上测定的常数值计算,则

$$K = (6.672\times10^{-11}\,\mathrm{m^3\cdot kg^{-1}\cdot s^{-2}})\cdot M$$
$$= (39860047\pm5)\times10^7\mathrm{m^3\cdot s^{-2}}.$$

环绕速度(circumaviation velocity) 亦称第一宇宙速度. 人造卫星运动理论中的速度概念. 即从地球表面发射人造卫星,使之环绕地球作轨道运动所需的最小速度(V_1),即

$$V_1 = \sqrt{\frac{K}{r}},$$

式中 K 是地心引力常数,r 是卫星离地心的距离. 若取 r 为地球赤道半径,即 6378.14km,$K=3.986\times10^{14}\mathrm{m^3\cdot s^{-2}}$,则 $V_1=7.905$ km/s.

第一宇宙速度(first cosmic velocity) 即“环绕速度”.

逃逸速度(escape velocity) 亦称第二宇宙速度. 人造卫星运动理论中的速度概念. 即从地球表面发射人造卫星,使之能沿抛物线轨道运动所需的最低速度. 具有此种速度的卫星可以逃离地球不再返回. 此速度为

$$V_2 = \sqrt{\frac{2K}{r}} = \sqrt{2}V_1 = 11.18\ \mathrm{km/s},$$

式中 K 是地心引力常数,r 是地球赤道半径,V_1 是第一宇宙速度.

第二宇宙速度(second cosmic velocity) 即“逃逸速度”.

第三宇宙速度(third cosmic velocity) 宇航学中的速度概念. 即从地球表面升空的飞行器,摆脱地球和太阳引力作用最终飞出太阳系所需的最低速度,即

$$V_3 = [V_2^2 + (V_{S_2} - V_E)^2]^{1/2},$$

式中 V_2 是第二宇宙速度,V_{S_2}是飞行器处于地球引力作用的轨道上欲沿抛物线脱离太阳引力作用所需

的日心坐标系中的速度，V_E 是地球绕太阳轨道的运动速度. 取 $V_2=11.18$ km/s，$V_{S_2}=42.12$ km/s，$V_E=29.77$ km/s，则 $V_3=16.66$ km/s.

第四宇宙速度(fourth cosmic velocity)　宇航学中的速度概念. 即从地球表面发射探测太阳的飞行器，使之飞达太阳所需的发射速度，即

$$V_4=(V_2^2+V_E^2)^{1/2}.$$

取 $V_2=11.18$ km/s，$V_E=29.77$ km/s，则 $V_4=31.8$ km/s(V_2，V_E 的意义见"第三宇宙速度").

轨道(orbit)　描述卫星运动的基本概念. 即天然或人造天体在引力中心体的引力作用下运动的路径. 在理想的中心力场作用下，轨道曲线是平面曲线，或为椭圆，或为抛物线，或为双曲线，随天体的相对位置和运动速度的不同而异. 但实际上天体运动并非只受中心体引力作用，周围其他天体的引力、中心体自身质量的非球形分布等因素，都会使天体轨道运动随时产生异常，轨道不会是平面曲线，而是像螺旋线那样的空间曲线. 通常，为了研究分析的方便，将这种复杂的轨道曲线分解为正常部分和异常部分. 其正常部分是瞬时的平面曲线，例如是椭圆. 异常部分是从一个瞬时椭圆到下一个瞬时椭圆的变化，这种变化包括椭圆参数，即形状大小的变化以及椭圆轨道平面在空间方向的变化. 轨道运动的这种异常又称为摄动. 有摄动的复杂轨道称为受摄轨道. 每一瞬时椭圆称为密切椭圆或密切轨道.

轨道根数(elements of orbit)　卫星轨道理论中的基本概念. 定义卫星轨道空间位置的参数. 即描述并确定卫星运动轨道的形状、大小、轨道面的空间位置以及卫星在其轨道上位置的参数. 以椭圆为例，共有六个根数：椭圆的半长轴 a；偏心率 e；椭圆轨道平面相对于地球赤道面的倾角 i；轨道椭圆与赤道大圆的交点之一升交点的经度 Ω；椭圆长轴与椭圆相交的两点(分别称为近地点和远地点的两点)中的近地点到升交点的角距，即近升距 ω；卫星在其椭圆轨道上飞经近地点的时刻 t_0. 这六个根数中，a，e 确定椭圆的形状大小，i，ω，Ω 确定轨道在空间的方位，t_0 确定卫星的轨道位置 .

近地点(perigee point)　见"轨道根数".

远地点(apogee point)　见"轨道根数".

密切轨道(osculating orbit)　亦称密切椭圆. 卫星轨道理论中的一个概念. 即一种逼近卫星实际轨道的无摄动轨道. 是卫星实际轨道运动中的一种人为假想的轨道. 是将有摄动力作用的受摄卫星运动轨道看做是一个个瞬时无摄动椭圆的合成轨道，每一无摄动椭圆与受摄轨道相切，切点处具有相同的卫星运动速度. 每一瞬时无摄动椭圆称为密切轨道(参见"轨道").

密切椭圆(osculating ellipses)　即"密切轨道".

同步轨道(synchronous orbit)　卫星轨道理论中的一个概念. 即一种接近于和地球自转同步运动的卫星轨道，或者说卫星绕地球一周的轨道周期和地球自转周期相同的卫星轨道. 这种卫星的轨道近于圆，半径约 36000km. 卫星绕地球作轨道运动时，相对于地面几乎没有相对运动，好像是静止地悬于空中，故称对地静止卫星. 洲际通信卫星、气象卫星以及 21 世纪计划发射的太阳能发电站等飞行体都是利用这种轨道特性.

太阳同步轨道(sun-synchronous orbit)　卫星轨道理论中的一个概念. 即一种轨道面相对地球赤道面的转动速度接近地球公转速度的卫星轨道；或者说，卫星运动时，轨道平面相对于太阳保持着固定不变角度的轨道. 具有这种轨道特性的卫星对地面观测时，可以始终在相同太阳光照角度的情况下进行. 资源遥感卫星、侦察卫星、气象卫星等都可利用这一轨道特性获取更有价值的资料. 实现太阳同步轨道要利用地球赤道隆起的非球形部分质量对卫星轨道的摄动作用，使卫星轨道面与赤道面的交线不断地在赤道面上转动，其转动速度等于地球绕太阳公转的速度，即每天转动 360°/365，转动的方向与地球公转方向相反.

卫星星下点(sub-satellite point)　卫星大地测量学中的一个概念. 即卫星的地心矢径与地面的交点. 卫星在轨道上绕地球运动时，矢径不断地随卫星运动，星下点也在地面上形成星下点轨迹. 由于地球不断自转，卫星轨道面也在缓慢地受摄转动，卫星绕地球一周形成的星下点轨迹不能闭合，各周的星下点轨迹不会重叠，类似于螺旋曲线.

卫星轨道寿命(orbital life-time of a satellite)　卫星轨道理论中的一个概念. 即人造卫星在其轨道上存在的时间. 寿命长短视卫星离地面的高低、卫星的截面积大小及卫星的质量大小等因素而定. 高度越高、截面积越小、质量越大，或质量截面比越大的卫星，寿命越长. 由于 200km 左右高度的圆轨道卫星运动在较密大气层中，大气阻力迅速使其轨道缩小、高度下降，进入更稠密大气层，几天或几十天就可能被烧毁.

原子时(atomic time)　一种计时系统. 即由原子钟按原子振荡频率计时的时间. 由于世界时依据天体运动而变化，存在着不均匀性和观测精度不够高的缺点，从 1967 年起，国际上开始将原子时作为计量时间的新系统. 原子时的基本单位是秒. 秒长的定义是：铯原子基态的两个超精细能级，在零磁场下跃迁辐射 9192631770 周所持续的时间. 原子时与天体运动完全无关.

协调世界时(coordinated universal time) 一种计时系统. 即以原子时为基准,同时兼顾世界时的计时系统,简称协调时. 时间单位采用原子时秒. 原子时秒长与世界时秒长有微小差别,而且会逐渐积累,使协调时和世界时指示的时刻相差越来越大. 为了彼此协调一致,特采取了一种"闰秒"的补偿办法,即当两者之差达到一定大小时,将协调时人为地增加或减少 1s. 跳秒调整的结果将使两者之差始终不超过±0.7s. 一般在每年 6 月 30 日或 12 月 31 日进行调整. 现在全世界各授时台播发的时间讯号大多是协调时时间讯号.

GPS 卫星时(GPS satellite time) 一种卫星计时系统. 即美国发射卫星启用的全球定位系统(GPS)专用的一种计时系统. 时间单位采用原子时秒,计时不用年月日,而是用离零时元的星期数,不足一星期的部分则用秒来累计,星期开始(即上星期六和星期日之间的子夜)取为 0s,到星期末为 604800s. 零时元是 1980 年 1 月 6 日协调世界时零时(记为 UTC 0^h). 由于 UTC 有跳秒,GPS 卫星时和 UTC 也不一致,两者之差可以从 GPS 卫星发播的电文中得知.

卫星大地测量学(satellite geodesy) 大地测量学的一个新分支. 是一门研究利用观测卫星轨道运动和信号进行大地测量的学科. 研究内容主要包括:地面点卫星定位、卫星重力探测以及用卫星技术监测地球形状和运动状态变化等的理论和方法.

卫星绕地球的运动,如果认为全部地球质量都集中于质心,则卫星绕质心运动的正常轨道服从牛顿引力定律,是一个椭圆. 但由于地球是个扁球,赤道隆起,内部质量分布密度不均匀、也不规则,周围空间有大气层和其他天体的引力作用,因此卫星并不是严格在中心力场中运动,运动轨迹不会是平面曲线,而可能是类似于空间螺旋线那样的复杂曲线. 通常是将这种复杂运动分解为正常运动和异常运动两部分. 正常运动部分是认为每个瞬时有一个平面曲线椭圆,可以用六个轨道根数来描述这个椭圆轨道的形状大小、轨道面相对于地球赤道面的位置以及卫星在此轨道上的起始位置. 异常运动部分则是指卫星的实际运动轨道与上述正常运动轨道的差异,常称为摄动,用轨道根数的变化量来表示. 计算卫星轨道,特别是计算摄动相当复杂,变量多、算式繁、数据量大的卫星轨道,要运用数学、天体力学、计算机科学等方面的知识.

如果以地球质心为原点建立一个三维直角坐标系,在此坐标系中,地面测站点 T 的位置矢量是 $\boldsymbol{r}_T$,在正常椭圆轨道上的卫星 S 的瞬时位置矢量是 $\boldsymbol{r}_S$,而地面站点对卫星的观测矢量此时是 $\boldsymbol{r}_{TS}$,三者的关系是

$$\boldsymbol{r}_T + \boldsymbol{r}_{TS} = \boldsymbol{r}_S.$$

式中 $\boldsymbol{r}_{TS}$ 总是可以用各种方法(例如光学观测、电磁波测距)测定的. 因此,在已知 $\boldsymbol{r}_{TS}$ 的条件下,如果要求测定地面测站点 T 的位置矢量 $\boldsymbol{r}_T$,还应该知道 $\boldsymbol{r}_S$. 反之,若是要测定 $\boldsymbol{r}_S$,则应该知道 $\boldsymbol{r}_T$. 在卫星大地测量应用中这两种情况都很实用,前者称为卫星定位,后者称为卫星定轨.

在卫星定位测量中,对卫星的测量或由单个测站进行,或由几个测站同时进行同步观测. 一系列地面测站点经卫星定位后便可构成地面的卫星定位网,用以满足各种建设对控制网的需要.

卫星绕地球运动的轨道异常主要由地球的扰动重力场产生,它反映了地球的物质分布不均匀和不规则. 由卫星轨道摄动观测数据可求解地球重力场,卫星测高技术可直接观测大地水准面的起伏,由此也可求解地球重力场,并精确确定地球形状;卫星技术还可测定地球自转参数及其变化,也是监测地球板块运动和地壳形变的有效手段. 卫星大地测量学是经典大地测量发展到现代大地测量的标志.

卫星定位(positioning by satellite observation) 见"卫星大地测量学".

卫星定轨(determination of satellite orbit) 见"卫星大地测量学".

卫星多普勒定位(satellite Doppler positioning) 卫星定位方法之一. 即测量卫星发射的信号载波频率对测站的多普勒频移由此确定测站位置的方法. 卫星绕地球质心运动,对于地面测站有相对运动,其运动速度可分解为沿轨道的分量和沿卫星到测站方向的径向分量. 径向分速度会影响卫星信号向测站传送的速度,从而产生多普勒频移. 若测量出一定时间间隔(例如 2min)内的平均频移量,就可算出在其起始和终止时刻卫星到测站的距离之差. 连续测出若干个距离差,以每一距离差所对应的卫星轨道位置为焦点作一双曲线,测站必在此双曲线上. 用三个以上的距离差做出三个以上的旋转双曲面,就可交会确定测站的位置.

20 世纪 70 年代,国际上普遍应用美国发射的子午卫星进行多普勒定位,并生产了专用接收机. 测量频移、进行电离层改正或加以抵消、对卫星广播电文解码、计算测站位置等工作,都可由接收机完成.

根据对成果精度要求的不同,多普勒定位可以由单个测站进行单点定位、由两个测站同时进行联测定位以及由多个测站同时进行短弧定位. 单点定位精度为几米,多用于海上导航定位;联测定位和短弧定位精度较高,可达 0.5m,可用于大地测量建立控制网.

全球定位系统(global positioning system) 简称 GPS. 卫星定位方法之一. 即利用 GPS 卫星发射

的信号确定地面点坐标的方法. GPS 卫星是美国继子午卫星系统之后发射的新一代导航卫星系统,共含 21 个卫星(另加 3 个备用卫星),分布于 6 个轨道,轨道面与地球赤道面的交角约为 55°,6 个轨道在赤道上均匀分开(彼此间隔 60°),每个轨道上各有 3～4 个卫星(也均匀分开). 因此,整个系统的卫星在空间基本上是均匀分布,任一地面测站随时都能接收到 4 个以上卫星的信号. 卫星轨道近于圆,平均半径为 26500km. 每个卫星都用两种频率的电磁波作为信号载频,分别是 1227.60MHz 和 1575.42MHz,目的是便于抵消电离层影响. 这一系统称为全球定位系统.

用 GPS 卫星进行测站定位的基本原理是距离交会. 测量电磁波在卫星与测站之间传播的时间,乘以光速常数,可以测定每一卫星到测站的距离. 用每一星站距离为半径,以卫星为球心,作一个球,若测定了三个星站距离,便可做出三个球,其交点便是待定的测站位置. 若多于三个星站距离,还可提高解的精度. 由于需要非常精确的时间,每个 GPS 卫星上都装有原子钟.

甚长基线干涉测量(very long baseline interferometry) 简称 VLBI. 一种空间大地测量方法. 即高精度测定地面上相隔很远的两测站点间空间距离和方向的现代空间技术. 其原理是测量电磁波干涉效应. 设在 A,B 两测站点上各安置接收电磁波的高性能天线,同时接收来自遥远空间某信号源 S(例如自然界的类星体或人造天体) 发出的同一频率电磁波. 如果信号先后到达 A,B 站的天线,有波程差,便会形成干涉. 当波程差恰等于信号波的半波长偶数倍时,信号加强;等于奇数倍时,信号减弱或抵消. 前者同相,后者异相. 因此,可以测量出同一信号波前经过此波程差的时间延迟量 τ. 于是相应的波程差便是 $c\tau$,其中 c 是光速.

设已知信号源 S 在天球坐标系中的位置矢量为 $\boldsymbol{S}\ (\alpha_S,\delta_S)$,其中 α_S 为 S 的赤经,δ_S 为 S 的赤纬. 待求定的是 $\overrightarrow{AB}(\Delta x,\Delta y,\Delta z)$,其中 Δx,Δy 和 Δz 是 A,B 两站点在地球空间直角坐标系中的坐标差,因为天球坐标系的赤道面与地球赤道面平行,这两种坐标系之间有确定的转换关系,因此观测的是 $c\tau$,三者的关系是

$$c\tau = -\ (\boldsymbol{S}\cdot\overrightarrow{AB}).$$

显然,只要观测三个 $c\tau$ 值,便可求出 $\overrightarrow{AB}$ 的三个分量,即求出 A,B 两站间基线 $\overrightarrow{AB}$ 的长度和方向. 实际上,每次观测可以获得大量的同步观测数据,还可用最小二乘法求解.

实现上述过程有很大的技术难度,关键是要准确测出 τ,这要将两站同步观测的记录进行相关处理,进行干涉比对,求出相关函数 $R(\tau,t)$,其中 t 为时间,再经计算机分离出干涉条纹的相位函数、振幅函数、时间延迟以及延迟率. 这对仪器设备的要求也很高,要有高性能接收天线、高稳定度原子钟,以及高灵敏度低噪声的接收机.

动力大地测量学(dynamic geodesy) 大地测量学的一个新分支. 用大地测量方法监测、研究地球动态变化的学科. 现代大地测量可精确测定地球整体运动、地面点位置和地球重力场要素随时间的变化,并研究这些变化和做出物理的解释. 动态大地测量中所测定的地球运动状态可分为三类,即地球重力的变化以及由此产生的大地水准面形状和垂线方向的变化;地球自转轴方向在空间的变化(岁差和章动)和在地球体内的变化(极移)以及地球自转速度的变化(日长变化);地球形变运动,它包括全球性板块运动和板块内的地壳运动以及潮汐引起的地球形变. 为了测定地球运动状态,需要采用多种高精度的测量手段. 除了用传统的大地测量方法外,还要采用新的空间大地测量手段. 前者包括高精度重复水准测量、天文测量、重力测量等,后者包括甚长基线干涉测量、卫星激光测距、卫星多普勒定位和 GPS 卫星定位、卫星雷达测高、卫星跟踪技术等.

实用天文学

实用天文学(practical astronomy) 天体测量学的一个分支学科. 它以球面天文学为基础,研究通过观测合适的天体以确定地面点的天文坐标和任一方向的天文方位角. 为此,需研究测定天文经纬度和天文方位角的原理和方法、观测仪器的使用、观测纲要的编制、测量结果的处理及误差的影响等问题.

早期的实用天文学根据不同部门的需要分为大地天文学、航海天文学和航空天文学. 随着近代无线电导航技术的发展,航海天文学和航空天文学已经逐步发展成为导航学中的一个分支学科——天文导航. 现代实用天文学为经济建设提供有关资料,并为地球物理学、地质学、地理学、地球动力学和制图学提供必要的参考数据.

中国实用天文学在历史上已有巨大成就. 根据天象变化推算四季规律并制定历法,已有悠久的历史. 公元 8 世纪初,唐代天文学家一行组织了中国第一次天文大地测量,用观测北极星高度法测定测站的天文纬度,由此推算出子午线 1°的弧长. 该成果为后人研究地球形状和大小以及纬度测量奠定了理论基础. 元代天文学家郭守敬制作了多种灵巧精密的天文测量仪器,并组织了全国规模的纬度测量工作.

大地天文定位常用瑞士威特厂生产的高精度 T_4 全能经纬仪进行. 采用太尔各特法测定天文纬

度，采用双星等高法并同时收录无线电时号测定天文经度，采用北极星任意时角法测定天文方位角，也可采用多星等高法同时测定天文经纬度．天文经度、纬度、方位角所能达到的精度分别高于±0ˢ．02，±0．30″和±0．50″．高精度天文经纬度可用来推算测站的天文大地垂线偏差，高精度天文方位角可以为一、二等三角测量和导线测量提供独立的拉普拉斯方位角控制．

天文点(astronomical point)　国家天文大地网中一类大地点．即用天文测量方法测定天文经度和天文纬度的大地点．

天文基本点(astronomical fundamental point)　亦称天文经度基本点．一种高精度的天文经度点．即在全国范围内按纬度相隔若干度并顾及地区分布，统一施测具有较高精度的天文经度点(精度高于±0．045″sec φ)．该天文基本点供大地天文观测员测定人仪差之用，使全国测定的经度纳入统一系统．中国在上海、北京、西安、武汉、广州、昆明、乌鲁木齐、拉萨、哈尔滨均施测有天文基本点．

天文经度基本点(astronomical longitudinal fundamental point)　即"天文基本点"．

天体(celestial body)　天文学基本名词．即宇宙间物质聚集形成的各种物体的通称，例如星云、恒星(含太阳)、行星(含地球、人造行星)、卫星(含月球、人造卫星)、小行星、彗星和流星等．

天轴(celestial axis)　描述天体视运动的一种轴线．即地球自转轴无限延伸与天球相交的轴线，也是天体在天空中每天东升西落环绕运行时的一根转轴．

天极(celestial pole)　天球上的一种特殊点．即天轴和天球相交的两点．与地北极对应的称为北天极，与地南极对应的称为南天极．

北天极(north celestial pole)　见"天极"．

南天极(south celestial pole)　见"天极"．

天球坐标系(celestial coordinate system)　表达天体位置的各种坐标系的总称．即确定天体在天球上的视位置而采用的球面坐标系．根据不同需要，天球坐标系由基圈、基圈的极、主圈以及主点和各坐标的量度方向来规定．在天文学中常用地平坐标系、赤道坐标系、黄道坐标系和银道坐标系四种．

天文子午面(astronomical meridian plane)　天球坐标系中的一个平面．即包含某一地面点的天顶及天轴的平面．

天文子午线(astronomical meridian line)　天球坐标系中的一种交线．即天文子午面与天球的交线．

天球赤道面(celestial equatorial plane)　见本卷《天文学》中的"天赤道"．

天球赤道(celestial equator)　见本卷《天文学》中的"天赤道"．

地平经圈(azimuth circle)　亦称垂直圈．地平坐标系中的一种大圆．它是地平坐标系的副圈．通过天球上天顶和天底的大圆称地平经圈．天体地平坐标系的高度或天顶距，就是在通过该天体的地平经圈上来量度的．

垂直圈(vertical circle)　即"地平经圈"．

地平纬圈(altitude circle)　亦称平行圈．地平坐标系中的一种平行圈．即天球上与地平圈平行的小圈．在同一地平纬圈上各点的高度(或天顶距)都相等，所以又称等高圈．

平行圈(parallel circle)　即"地平纬圈"．

等高圈(almucantar)　即"地平纬圈"．

赤经圈(circle of right ascension)　亦称时圈．赤道坐标系中的一种大圆．即通过天极的大圆，它是赤道坐标系的副圈．赤经圈有无穷多个，都与天球赤道垂直．

时圈(hour circle)　即"赤经圈"．

天体视运动(apparent motion of celestial bodies)　一种自然现象．即人们在运动着的地球上所看到的天体的运动现象．由于地球自转，地球上的观测者会看到：天体于一恒星日内，在天球上自东向西沿着与赤道平行的小圈完成一周运动．这种直观的天体运动称为天体周日视运动．周日视运动是一切天体最显著的视运动．天体在作周日视运动时，经过天球上一些特殊的圈时有特定的名称，在天体测量中具有重要意义．

天体经过观测者的子午圈时为中天；经过包含天极和天顶那半个子午圈时，天体到达最高位置，称为上中天；经过包含天极和天底那半个子午圈时，天体到达最低位置，称为下中天．天体经过观测者的地平圈时称出没，也称升落．天体从地平圈之下升到地平圈时称为出，反之称为没．

由于地球公转，人们看到的星座在一年四季内是不同的，但每年重复一次，称为天体周年视运动．

上中天(upper culmination)　见"天体视运动"．

下中天(lower culmination)　见"天体视运动"．

天体周年视运动(annual apparent motion of celestial bodies)　见"天体视运动"．

周日平行圈(diurnal parallel circle)　赤道坐标系中的一种平行圈．即天体作周日视运动时，在天球上绕天轴旋转一周的轨迹．它是平行于天球赤道的小圈，该小圈即为天体的赤纬圈．地球每自转一周，天体就沿其周日平行圈运行一周．

拱极星(circumpolar star)　在天球上一种处于

特殊位置的恒星.即整个周日视运动轨迹都在地平面以上的恒星.它们的赤纬需满足$\delta \geqslant 90° - \varphi$的条件.由于它们位于北极附近,周日视运动的线速度较小,故可作为一些特殊天文测量的观测对象.天体测量习惯上把赤纬大于80°的近极星称为拱极星.

大距(greatest elongation)　对天体在运行中处于特定位置时的命名.

1.在地球上观测内行星和太阳之间的角距称为距角.距角达最大时为大距,在太阳之东称为东大距,在太阳之西称为西大距.水星的大距为28°,金星的大距为48°.

2.天体在其周日平行圈上运行,当通过天体的地平经圈与周日平行圈相切时,天体与北极之间的距角达到最大,称为大距.满足$\delta > \varphi$条件的天体才会有大距.有大距的天体,在一恒星日内有两次大距,一次发生在东边称为东大距,一次发生在西边称为西大距.大距时天体的星位角$q = 90°$,其方位角取得极值,此时天体只作垂直运动而无水平运动,即天体在大距附近时其方位角的变化率最小,有利于天文方位角的测定.

距角(elongation)　见"大距".

夏至点(summer solstice)　决定地球四季转换的天文分界点之一.即太阳在黄道上周年视运行时所经过的四个特定点之一.黄经为90°的一点为夏至点,亦即每年6月22日前后,视太阳到达距赤道最北的那一点.这一天北半球白昼最长,天文学上规定夏至为北半球夏季的开始.

秋分点(autumnal equinox)　决定地球四季转换的天文分界点之一.即太阳在黄道上周年视运行时所经过的四个特定点之一.天球上黄道和赤道相交的两点之一,每年9月23日前后,视太阳沿黄道由赤道以北通向赤道以南时的一点称为秋分点.天文学上规定秋分为北半球秋季的开始.

冬至点(winter solstice)　决定地球四季转换的天文分界点之一.即太阳在黄道上周年视运行时所经过的四个特定点之一.黄道上黄经为270°的一点为冬至点,亦即每年12月22日前后,视太阳到达距赤道最南的那一点.这一天北半球白昼最短,天文学上规定冬至为北半球冬季的开始.

近日点(perihelion)　太阳系天体在轨道上公转时所经过的两个特定点之一.行星、小行星、周期彗星以及人造行星等环绕太阳公转的轨道上离太阳最近的一点.天体到达近日点时公转速度最快.地球近日点距太阳1.47099×10^8km.

远日点(aphelion)　太阳系天体在轨道上公转时所经过的两个特定点之一.行星、小行星、周期彗星以及人造行星等环绕太阳公转的轨道上离太阳最远的一点.天体到达远日点时公转速度最慢.地球远日点距太阳1.52101×10^8km.

年(year)　一种天文计时单位.即以地球绕太阳公转运动为基础的时间单位.地球公转运动的周期是借天球上太阳的周年视运动来反映的.在天球上选择不同的参考点,就会得到不同的太阳周年视运动的周期,亦即能定义出不同的年,以供不同的需要.

以春分点为参考点定义的年称为回归年,每年等于365.24220平太阳日;以恒星为参考点定义的年称为恒星年,每年等于365.25636平太阳日;以近日点为参考点定义的年称为近点年,每年等于365.25964平太阳日;以月球轨道升交点为参考点定义的年称为交点年,每年等于346.62003平太阳日.

回归年(tropical year)　见"年".

恒星年(sidereal year)　见"年".

近点年(anomalistic year)　见"年".

交点年(nodical year)　见"年".

月(month)　一种天文计时单位.即以月球绕地球公转运动为基础的时间单位.在天球上选择不同的参考点或不同的起算点,就会得到不同的月.以月相变化的周期定义的月称为朔望月,每月等于29.53059平太阳日;以天球上月球连续两次通过某一恒星所定义的月称为恒星月,每月等于27.32166平太阳日,这是月球的平均公转周期;以月球黄经连续两次等于春分点黄经所定义的月称为分点月,亦称回归月,每月等于27.32158平太阳日;以月球连续两次经过近地点所定义的月称为近点月,每月等于27.55455平太阳日;以天球上月球连续两次由南向北通过黄道升交点所定义的月称为交点月,每月等于27.21222平太阳日.

朔望月(synodic month)　见"月".

恒星月(sidereal month)　见"月".

回归月(tropical month)　见"月".

近点月(anomalistic month)　见"月".

交点月(nodical month)　见"月".

时间(time)　物质存在的基本形式和基本物理量之一.物质和运动是不可分割的,物质运动的存在,由时间和空间来体现,时间和空间是物质的两个属性.时间可以通过某种选定的物质运动过程(必须是连续的、均匀的、可以赋予或获得的)来计量.通常所说的时间计量,实际上包含既有差别又有联系的两个方面:时刻和时间间隔的计量.时间间隔(简称"时间")是指客观物质运动在两个不同状态瞬间所经历的时间历程.时刻是指客观物质在其某一运动状态瞬间与时间坐标轴原点之间的时间间隔,是没有延续性的瞬间.从广义上说,目前能计量的时间间隔,可从天体的年龄到极短的物理过程,幅度范围在

$10^{-24}\sim10^{18}$时秒之间.天文学上的时间计量是专指一"日"以内的时间间隔计量,以及选定初始历元作为时间计量的起点.

真太阳日(true solar day) 亦称视太阳日.一种天文计时单位.即以地球自转为基础的一种时间单位.是真太阳中心连续两次上中天经历的时间长度.由于真太阳在黄道上周年视运动速度的不均匀,以及黄道和赤道具有交角,所以真太阳日的长度也不均匀,即是一个不均匀的时间系统.一年中最长和最短的真太阳日约差 51 时秒,1 真太阳日分为 24 真太阳时,1 真太阳时分为 60 真太阳时分,1 真太阳时分分为 60 真太阳时秒.

视太阳日(apparent solar day) 即"真太阳日".

真太阳时(true solar time) 亦称视太阳时或视时.一种天文计时系统.视时等于真太阳中心的时角加 12 时,以真太阳中心下中天瞬间为视时的开始(即零时).视时从下子午圈起计量,不同经度的观测者在同一瞬间的视时是不同的,以地方视时来表示.

视太阳时(apparent solar time) 即"真太阳时".

视时(apparent time) 即"真太阳时".

时差(equation of time) 两种计时系统之间的时间差.即任一瞬间真(视)太阳时与平太阳时的差.常定义为真(视)太阳时减平太阳时,用 E 表示.中国天文年历太阳表中列出每天 0^hTDT(力学时 0^h)的时差 E.一年中时差在−14 分 24 秒至+16 分 21 秒之间变化,它的周日变化值为−30 秒至+21 秒.时差值每年在 4 月 16 日、6 月 15 日、9 月 1 日和 12 月 24 日前后有四次等于零;还有四次为极值:2 月 12 日前后为−14.4 分,5 月 15 日前后为+3.8 分,7 月 26 日前后为−6.3 分,11 月 3 日前后为+16.4 分.

本初子午线(prime meridian or first meridian) 亦称首子午线或零子午线.地球上经度的起算经线.19 世纪以前,许多国家采用通过大西洋加那利群岛耶罗岛的子午线(西经 17°39′46″)为本初子午线.19 世纪上半叶,改用通过本国主要天文台的子午线为本初子午线.1884 年,国际经度会议决定,以通过格林尼治天文台埃里中星仪的经线为本初子午线,全球经度以它作为起始经线.1957 年后改用若干测时长期稳定性较好的天文台组成的平均天文台作为经度零点.1968 年,以国际协议原点(CIO)作为地极原点,把通过国际协议原点和平均天文台经度零点的子午线作为本初子午线.

首子午线(prime meridian) 即"本初子午线".

零子午线(zero meridian) 即"本初子午线".

世界时系统(universal time system) 一种计时系统.即以地球自转为基础的时间计量系统.为了测量地球自转,人们在天球上选取两个基本参考点:春分点和平太阳,由此确定的时间分别称为恒星时和平太阳时.通常通过观测恒星求得恒星时,再算得平太阳时.

1884 年,以平子夜作为零时开始的格林尼治平太阳时称为世界时,以 UT 表示,亦称格林尼治平时或地球自转时.

1955 年,第九届国际天文学联合会通过了世界时系统的定义:

1. 由天文观测直接测定的世界时系统称为 UT0 系统.长期以来,UT0 系统被认为是均匀的时间计量系统,曾得到广泛应用.

2. 在 UT0 系统中引入极移改正 $\Delta\lambda$ 的世界时系统称为 UT1 系统.

3. 在 UT1 系统中引入地球自转速率季节性变化 ΔT_s 的世界时系统称为 UT2 系统.

它们之间的关系是

$$\begin{aligned}\mathrm{UT1} &= \mathrm{UT0} + \Delta\lambda,\\ \mathrm{UT2} &= \mathrm{UT1} + \Delta T_s\\ &= \mathrm{UT0} + \Delta\lambda + \Delta T_s.\end{aligned}$$

恒星时(sidereal time) 见"世界时系统".

平太阳时(mean solar time) 见"世界时系统".

格林尼治平时(Greenwich mean time) 见"世界时系统".

地球自转时(earth rotation time) 见"世界时系统".

北京时刻(Beijing time) 中国通用的标准时刻,即东经 120°经线为准的东八时区区时.北京的经度是东经 116°21′,因此北京时刻并不是北京地方平时,北京时刻比北京地方平时约早 14 分 30 秒,比格林尼治平时(世界时)早 8 小时.例如,10 月 1 日北京时刻 8 时,这时世界时为 10 月 1 日 0 时.

平均天文台(mean observatory) 用数学方法算得虚设的零经度的天文台.即选择若干观测时间长、稳定性较好的天文台站,用天文测时资料反求这些台站经度采用值的改正数,即求得每个台站各自的经度零点,再对这些经度零点统一进行平差处理,所求的最终平均经度零点称为平均天文台.

地理坐标(geographic coordinates) 一种表示地理位置的坐标.即指地面点位置用经纬度表示的球面坐标.一般系指大地坐标,即大地经度和大地纬度,有时也指天文经度和天文纬度,两者的差为垂线偏差的影响,平原地区一般为几角秒,山区通常不超过 40″.对于小比例尺地形图的点位来说,两种地理坐标可不加区别,但在大地测量中正是利用这两种坐标之差计算天文大地垂线偏差.

天文经度(astronomical longitude)　量度地球表面东西方向间距的量. 即地面点的天文子午面与本初子午面间的二面角,或这两子午面在赤道上所夹的弧段,以 λ 表示. 从本初子午面分别向东、向西量度,各自为 0～12 时或 0°～180°,在本初子午面以东称为东经,以西称为西经.

天文纬度(astronomical latitude)　量度地球表面南北方向间距的量. 即地面点的铅垂线与天球赤道面的交角,以 φ 表示. 从天球赤道分别向南北两极计量,各自为 0°～90°,在天球赤道以北取正号,称为北纬;以南取负号,称为南纬.

天文方位角(astronomical azimuth)　量度地面目标方向线偏离正北的角度. 即从地面上某点的天文子午线指北方向,按顺时针方向至另一目标方向线间的水平角,在实用天文学中常用 A_N 表示,自 0°至 360°.

日界线(date line)　亦称国际日期变更线. 地球上的日期分界线. 1884 年,国际子午线会议决定采用 180°经线为日界线,地球上的日期以此日界线为界,两侧的日期不同,由线东向线西过日界线(从美洲到亚洲),日期要增加一天(即为第二天的日期). 反之(从亚洲到美洲),日期要减少一天(即为前一天的日期). 考虑到日界线所经过的国家和地区使用两个日期的不便,实际的日界线是从北极开始,沿 180°经线前进,曲折通过白令海峡,绕过阿留申群岛西边,穿过萨摩亚、斐济、汤加等群岛之后,再经过新西兰东边,最后又继续沿着 180°经线终止于南极.

国际日期变更线(international date line)　即"日界线".

时间服务(time service)　一种时间信息服务系统. 即为科学研究和生产提供标准时刻和频率的信息服务. 它是靠无线电、电视等手段来服务的. 总的来说,时间服务工作包括测时、守时、播时、计算、出版时号改正数等. 中国上海天文台和陕西天文台等单位是主要从事时间服务的科学研究机构.

时号(time signal)　一种时间信息服务信号. 即用无线电发播的时刻信号. 授时工作的天文台将测定的精确时刻,以一定的信号及程序,每日若干次或连续地用无线电发播出时号,供天文、测绘、宇航、无线电等计量部门应用. 按时号的型式分,有平时式时号、国际式时号和协调世界时时号等. 中国天文台通过中央人民广播电台,在每个整点发播的六个声响也是时号.

协调世界时时号(UTC time signal)　一种时间信息服务信号. 即根据协调世界时播发的无线电时号. 目前,世界上各授时天文台发播的时号,大多是协调世界时时号. 中国陕西天文台也发播协调世界时时号 BPM_c.

时号改正数(correction to time signal)　一种对时间信息服务信号的改正. 即每次发播时号所代表的计划时刻与实际发播的准确时刻的差数. 时号所代表的计划时刻加上时号改正数,即等于时号所代表的实际准确时刻.

综合时号改正数(synthetic correction to time signal)　一种时间信息服务信号的改正. 即综合处理多个天文台站的测时结果所得到的时号改正数. 单独一个台站的测时结果,由于受到气候条件的限制,多半时断时续,全年分布不均匀;又因地区性反常折射和人仪差等影响,使求得的时号改正数的精度受到很大的影响. 利用全世界、全国或一定区域有关台站的测时结果,经过处理归算后得到的综合时号改正数,其精度可以大大提高.

中国的综合时号改正数是由上海天文台综合处理归算出版的,其精度始终列居世界先进水平.

授时公报(time service bulletin)　一种时间信息服务刊物. 这里指中国时间服务的定期刊物. 其内容从前是北极的瞬时坐标(x,y)、地球自转季节性变化经验改正 ΔT_s(用于 UT2 系统的归算)和以中国世界时综合系统(UT1)为准的综合时号改正数. 自从采用原子时后,国际上无线电时号一般均按协调世界时或原子时发播. 1980 年 1 月起,中国授时公报中不再给出各种时号的改正数,而改为提供每天世界时(UT1)和协调世界时(UTC)、原子时(AT)、中国世界时时号 BPV_1 及 BPV_3 的精确差值:

$$UT1-UTC,\quad UT1-AT,$$
$$UT1-BPV_1,\quad UT1-BPV_3,$$

同时提供中国协调世界时时号 BPV_2 相对 UTC 的发播值 $UTC-BPV_2$. 用户对所收录的时号根据其类型加上相应的差值,即可求得相应的精确世界时时刻. 国际上凡有时间服务的天文台,都按期出版授时公报,主要内容与中国授时公报相同.

地球自转参数公报(earth rotation parameters bulletin)　一种监测地球自转运动数据服务刊物. 这里指中国地球自转服务的定期刊物. 1984 年起,中国将授时公报改名为地球自转参数公报,内容未变. 1988 年 1 月 1 日起,国际上成立了国际地球自转服务(IERS),代替国际时间局(BIH)和国际极移服务(IPMS)等机构的工作. 新的机构利用甚长基线干涉测量、激光测卫、激光测月等新技术的观测资料,逐步淘汰了以经典光学仪器为主的观测资料,以确定地球自转参数. 地球自转参数公报的内容为:极坐标、世界时、岁差、章动、动力春分点、参考台站位置和指定射电源位置,以及地壳形变参数.

经度起算点(origin of longitude)　一种区域性经度基准点. 全球经度起算点以本初子午线为准,但

距离本初子午线较远的国家和地区，其经度就不便用本初子午线直接比对测定. 尤其是1968年以来，本初子午线有新的含义，因此有些国家在本国测定一个高精度的天文经度点，然后再换算到本初子午线上去. 1978年起，中国的经度起算点是上海天文台目视等高仪的中心，其经度值（以中国采用的极原点$JYD_{1968.0}$为准）为8时05分42.5067秒±0.0017秒.

瞬时极（instantaneous pole）　地球南北极点的一种定义. 即瞬时自转轴与地球表面相交的点. 地球自转时，地球本体围绕通过地球质心的自转轴旋转. 地球自转轴在地球体内并不是固定的，随时都有变化，地球自转轴在某一时刻的实际位置称瞬时自转轴. 因此，瞬时极在地面上的位置也随时间变化.

平极（mean pole）　地球南北极点的一种定义. 根据几个纬度观测站在某一时期内的平均纬度所确定的平均极. 它是用来计量极移和瞬时极坐标的参考原点.

固定平极（fixed mean pole）　平极的一种. 它是由若干个测站的固定平纬确定而得的平极. 所谓固定平纬是将每个测站在某一时期（一般为六年）内的纬度平均值作为该测站长期采用的平均纬度，是一个常数. 根据若干个测站的固定平纬推算而得的固定平极，认为是地面上没有漂移的固定点，如国际协议原点（CIO）.

历元平极（mean pole of the epoch）　平极的一种. 一定历元的瞬时极消除周期运动后的极点. 它是由一个或若干个测站以某一历元的纬度（不包含周期性变化）作为该历元的平均纬度（称为历元平纬）来推算而得的平极. 不同历元的历元平纬和历元平极均有微小的不同.

历元平纬（mean latitude of the epoch）　见“历元平极”.

国际协议原点（conventional international origin）　简称CIO. 地球北极点的一种国际上共同采用的定义. 即国际间采用的统一的地极坐标原点. 为了统一全球地面坐标系统和便于研究极移，1967年，国际天文学联合会（IAU）、国际大地测量学和地球物理学联合会（IUGG）决定采用统一的地极坐标原点，称为国际协议原点. 该原点是根据五个国际纬度站的平纬来确定的. 这些平纬是指归算到均匀系统后的1900—1905年期间的地球自转极的平均位置.

CIO确定了地球参考坐标系的主轴方向和相应的赤道面. 由于地壳运动，天文台按照某些规定方法建立的参考坐标原点的稳定性问题就显得非常突出.

极原点（JYD origin）　简称JYD. 中国采用的一种地极坐标系统. 中国搜集了国内外观测质量好、成果比较连续的41架仪器的测纬资料，运用滤波方法求得各测纬仪器所测定的纬度长期变化，进而解出极移的长期分量（x_0,y_0），而周期分量（x_1,y_1）则按单站法用中国的测纬资料计算得出. 原点采用1968.0年的平极，这是中国建立的地极坐标系统. 用极原点三字汉语拼音的缩写JYD加注历元1968.0——$JYD_{1968.0}$表示之.

天文定位测量（astrofix measurement）　一种定位技术. 即用天文测量方法测定地面点位地理坐标的技术. 在地面测站上用天文测量仪器测定天体的瞬时位置（垂直角和水平角，并记录相应的时刻），推算测站的天文经度λ、天文纬度φ和测站点至某照准点的天文方位角A_N，该过程称为天文定位测量. 用天文定位测量测定地面控制点，各点之间不受自然地理条件限制，不需要相互通视. 但天文观测的点位精度不高，除边远布网困难地区可用来作为1：100000或更小比例尺测图控制外，不能用它单独建立国家水平控制网. 但在建立国家水平控制网时，必须要有天文定位测量相配合，以进行参考椭球定位（参见“参考椭球定位”）；在天文大地网中用天文定位测量测定拉普拉斯点，还可控制大地测量中水平角观测误差的累积；天文重力水准测量可提供研究地球形状和大小的重要资料.

双星等高测时法（method of time determination by two stars in equal-altitude）　亦称东西星等高法或金格尔星对测时法. 一种测时方法. 即大地天文学中一种精密测时的方法. 其原理是在东西卯酉圈附近观测两颗恒星在等高h的时刻，以求钟（表）差u. 观测第一颗星（α_1,δ_1,t_1）的钟面时为T_1，观测第二颗星（α_2,δ_2,t_2）的钟面时为T_2，则可写出：

第一颗星 $\sin h=\sin\varphi\sin\delta_1+\cos\varphi\cos\delta_1\cos t_1$，

第二颗星 $\sin h=\sin\varphi\sin\delta_2+\cos\varphi\cos\delta_2\cos t_2$，

式中$t=S-\alpha=T+u-\alpha$，t为时角，S为地方恒星时. 合并两式消去$\sin h$后，即得解算u的方程. 由于两星等高，且计算式中用不到h，所以此法不受竖盘刻画误差和蒙气差误差的影响，从而大大地提高了测时精度.

东西星等高法（equal-altitude method with east and west stars）　即“双星等高测时法”.

金格尔星对测时法（method of time determination by Zinger star-pair）　即“双星等高测时法”.

北极星任意时角法（method by hour angle of polaris）　实用天文学和大地天文学中测定天文方位角的一种方法. 应用天文全能经纬仪测量北极星方向和地面目标方向间的水平角Q，并记下照准北极星瞬间的钟面时刻T，按下式求得北极星方位角a_N：

$$\tan a_{\mathrm{N}}=-\frac{\tan\delta\sec\varphi\sin(T+u-\alpha)}{1-\cot\delta\tan\varphi\cos(T+u-\alpha)},$$

式中 α,δ 分别为北极星赤经和赤纬，u 为钟差，φ 为测站纬度. 因为以上诸值均为已知，故可求得观测瞬间北极星的方位角 a_{N}. 再由北极星方向和地面目标方向间的水平角 Q，即可求得测站至地面目标方向间的天文方位角为

$$A_{\mathrm{N}}=a_{\mathrm{N}}+Q.$$

方位角下标 N 表示由北起算.

摄影测量与遥感学

摄影测量学(photogrammetry)　测绘学的重要分支. 它是通过摄影相片和辐射能的各种图象记录手段，在不接触所摄物体的情况下，对所摄目标的影像进行处理、量测、判释和研究，以测得被研究对象的形状大小和位置的图解形式或数字形式的成果，并获得关于环境等诸方面可靠信息的一门科学. 研究的主要对象是地球表面，用以测绘地形图. 随着遥感技术和各种光电传感器的发展，摄影方式和研究对象日趋多样，摄影测量已在许多领域广泛应用. 一般摄影均在可见光或近红外范围内，但也可从电磁波的其他波段中得到图象. 只要能得到图象就可以进行量测，因此对固体、液体、气体、宏观的、微观的、运动中的或变化着的对象均可应用.

摄影测量(photogrammetrical survey)　测量的一种重要方法. 利用摄影相片或其他方式的图象测定物体的形状、大小和空间位置的技术. 主要用于测绘地形图. 按相片和图象获得方式的不同分为航天摄影测量、航空摄影测量、地面摄影测量、近景摄影测量(又称非地形摄影测量)和水下摄影测量(又称双介质摄影测量)等.

地面摄影测量(terrestrial photogrammetry)　摄影测量的一种. 即利用摄影经纬仪在两个地面站摄取同一地段的立体像对，同时测定摄站点和若干相片控制点的坐标，并按相片进行地物调绘，然后在室内借助仪器测绘地形图. 主要用于测绘局部山区或困难地区的地形图及地质勘探中的地形地质测图，也可为各种工程中变形观测等提供数据.

航空摄影测量(aerophotogrammetry)　摄影测量的一种. 利用航空摄影相片测制地形图和获得其他信息的技术. 是由装在飞机或其他航空飞行器上的摄影机，按一定的规格从空中向地表摄影，取得相片后，通过室外的控制测量和室内加工处理等步骤，获得各种信息资料和测绘地形图. 外业工作包括相片联测、相片调绘和相片测图等；内业工作包括测图控制点加密、利用各类光学机械的立体测图仪或电子计算机辅助的测图仪器获取各种信息资料和测绘地形图. 若选用适当的航摄比例尺和测图方法，可测制各种比例尺地形图. 所得成果为线划图或影像地图以及数字地面模型(DTM)等. 根据成图过程，航空摄影测量分为综合法测图、模拟法测图(包括微分法和全能法)、解析测图和数字测图等. 航空摄影测量已在各个领域内得到广泛应用.

航天摄影测量(space photogrammetry)　摄影测量的一个重要分支. 是从人造地球卫星、航天飞机或宇宙飞船等航天飞行器上对被研究对象(主要是地球表面)用各种遥感器(包括相机)进行遥感，根据所获照片、图象和信息进行几何处理、分析、判释和测量，得到地形图和其他各种信息，统称航天摄影测量或卫星摄影侧量. 它主要应用于地球资源勘查和军事侦察、环保监测、测绘小比例尺地形图、月球图或其他天体图.

卫星摄影测量(satellite photogrammetry)　见“航天摄影测量”.

航空摄影(aerophotography)　对地面进行摄影的一种方式. 在飞机等航空飞行器上安装航空摄影机，从空中对地球表面进行的摄影. 航空摄影相片主要用于摄影测量，也用于农业、林业、地质、石油、水利、城市规划、环保和铁路、公路等线路的勘察以及军事等各个方面. 根据摄影方式的不同，可分为近似垂直摄影和倾斜摄影. 从卫星和航天飞机等航天飞行器上对地球表面的摄影称为航天摄影.

影像质量评定(quality evaluation of aerophotographic image)　航空摄影测量中的一个关键步骤. 即评定航空摄影相片的影像质量. 评定内容包括灰雾值、最大密度和最小密度、影像反差、反差系数、压平精度和各种标志(框标、气泡、时表等)的影像是否清晰. 其中灰雾值、最大密度和最小密度、影像反差和反差系数关系到负片的透光能力、影纹的显露和能否判识与精确量测，故对它们都有一定的数据要求；压平精度涉及影像位移，对它有严格要求；各种标志分别为确定像平面坐标系、检查相片倾斜角大小的依据和摄影时间的记录，均需明显呈现.

航摄相片(aerial-photo)　航空摄影的产品. 即用航空摄影机从空中向地面进行一定规格要求的摄影所取得的相片. 是航空摄影测量和航空地质等工作的基础资料，也是解决农林、水利和军事等方面有关问题的重要资料. 按航空摄影机在摄影时主光轴偏离铅垂方向的倾角，可分为竖直摄影相片和倾斜摄影相片两种. 倾角不大于 3°时所摄得的相片，称为竖直摄影相片. 按摄影机结构不同，又有画幅式相片、全景相片和航带连续摄影相片之分，最常用的是画幅式相片.

立体像对(stereopair)　简称像对. 从两个不同位置对同一地区所摄取的一对相片. 用立体观测法

和专用的工具可以在像对重叠影像部分内看出所摄目标的立体视模型.

立体摄影测量(stereophotogrammetry) 摄影测量的一种方法.利用立体像对,通过模拟的或解析的立体测图仪器,测定所摄物体的形状、大小和空间位置,以绘制地形图或等值线图的方法.航空摄影测量中的微分法测图、全能法测图和空中三角测量以及地面立体摄影测量,都属于这个范畴.

解析摄影测量(analytical photogrammetry) 摄影测量的一种.是用数学解算的方法进行测点、测图或空中三角测量的技术.数学基础包括:

1. 表达投影光线的像点、投影中心和地物点三点共线的条件方程式.
2. 表达像对中同名光线相交以建立立体模型的共面条件方程式.
3. 表达立体模型绝对定向的模型点空间坐标相似变换的关系式.

在应用中,都需将上述关系式线性化.在有多余观测的条件下,用最小二乘法进行计算.随着电子计算机的发展、软件功能的不断增强、解析测图仪的完善与普及,解析空中三角测量已应用于各个领域,成为摄影测量发展的一个重要方面.

数字摄影测量(digital photogrammetry) 20世纪80年代开始发展的摄影测量学的一个分支.航摄相片上的影像信息通过采样和量化,转变为大量的密集的灰度数字阵列,记录和存储在磁介质中,即转化为一张数字相片;或者采用新型的电荷耦合器件(CCD)阵列的摄影机,得到一幅光电成像的相片,记录和存储在磁介质中.然后,直接利用这些数字相片,通过计算机的控制和运算,进行找点、相片定向和测图等处理过程,并在专门的设备上自动晒印出带有等高线的正射影像地图.

相片比例尺(photo scale) 航摄相片的一种特性.即航摄相片上某线段长度与实地相应长度之比.当航摄相片严格水平时(相片倾角等于零),地面也是水平的理想条件下,整张相片才有统一的比例尺.但这种情况是不存在的.故航摄相片上的影像比例尺随着线段位置不同和地面起伏情况而变化.通常指的相片比例尺都是近似值.

相片重叠(photo overlap) 一种摄影测量的前提条件.指在摄影测量中,相邻相片上包含同一地区影像的部分.重叠的多少以占全张相片面积的百分数来表示,称为重叠度.足够的相片重叠是进行立体观测和模型连结所必须的.在航空摄影测量中,沿同一航线各相片间的重叠,称为航向重叠或纵向重叠;相邻航线间相片的重叠,称为旁向重叠或横向重叠.

重叠度(overlap) 见"相片重叠".

航向重叠(forward overlap) 见"相片重叠".

旁向重叠(lateral overlap) 见"相片重叠".

相片的内方位元素(elements of photo interior orientation) 摄影测量中的重要参数.包括相片主距和像主点坐标.亦即在摄影测量中,需确定摄影机物镜后节点相对于相片面的三个参数:相片主距 f,即物镜后节点到像主点的距离;像主点在相片框标坐标系中的 x 坐标值 x_0;像主点在相片框标坐标系中的 y 坐标值 y_0.通过摄影机检验得知,这三个参数均为已知值.

恢复光束测图(mapping by recovery of bundle of rays) 一种航摄测图方法.即保持相片内方位元素的投影光束测图.若测图时投影器与相片内方位元素不一致,则称为变换光束测图.

变换光束测图(mapping by rays bundle transform) 见"恢复光束测图".

相片的外方位元素(elements of photo exterior orientation) 摄影测量中的重要参数.即确定摄影瞬间相片与地面相对关系的参数.相片的外方位元素有六个:摄影物镜前节点在地面坐标系中的三个空间坐标值 x_s,y_s,z_s(三个直线元素);确定相片在地面坐标系中空间方位的三个角值:相片旋角 κ、倾角 α 和主垂面方位角 A.相片的外方位元素通常是未知数,需根据地面控制点反求.地面摄影测量时,由地面控制测量确定摄站坐标,并可按一定方式确定摄影方向,这时,相片的外方位元素乃为已知数.

像对的相对定向(relative orientation of photo pair) 处理摄影像对的一个步骤.即恢复或确定像对在摄影时的相对关系.理论根据是同名光线共面的原则.表示方式有两种:

1. 以摄影基线为基准,用像对两个光线束的五个旋转角值表示,称为单独像对相对定向;
2. 以其中一张相片为基准,用另一张相片的两个直线移动值和三个旋转角值表示,称为连续像对相对定向.

求解相对定向元素的方法也有两种:

1. 光学机械法.是在立体测图仪上利用投影器的旋转角和基线支架的直线运动,消除像对内(至少五个)同名点的上下视差.
2. 解析法.是在坐标量测仪上量测像对中(至少五个)同名点的像平面坐标,根据相对定向元素与上下视差的关系式计算相对定向元素值.完成像对的相对定向,就可认为已获得自由比例尺和处于空间任意位置的所摄地面的几何模型.

单独像对相对定向(relative orientation of single photo pair) 见"像对的相对定向".

连续像对相对定向(relative orientation of successive photo pair) 见"像对的相对定向".

模型的绝对定向(absolute orientation of mo-

del） 亦称几何模型的绝对定向. 确定像对在地面坐标系中位置的过程. 即确定由像对相对定向所获得的几何模型的三个旋转角值 $\bar{\Phi},\bar{\omega},\bar{K}$ 和其坐标原点的三个平移值 x_0,y_0,z_0 以及归化几何模型的比例尺，使它正确地置于地面坐标系中. 必须依据至少三个不在一条直线上的地面控制点，使模型上的相应点与之联系，求出模型的绝对定向的七个元素值. 求解方法有两种：

1. 模拟法. 像对在立体测图仪上完成相对定向后，根据三个以上控制点，改变投影基线的长度以归化模型比例尺，并使整个模型相对于承影面在两个互为垂直的方向上倾斜以置平模型，同时用手推移和旋转图底来达到绝对定向.

2. 解析法. 根据三个地面控制点的坐标值及其相应点在几何模型上的坐标值，采用三维坐标相似变换的数学运算，求解出上述七个定向元素值. 在区域网空中三角测量中，区域整体模型的绝对定向都采用解析法.

空间后方交会(space resection) 摄影测量中同时求解相片内、外方位元素的一种方法. 用坐标量测仪量测若干物方已知点（即地面坐标已知的点）的像点坐标，根据像点、物点和投影中心三点共线的关系式，在有多余观测条件下进行平差计算，确定相片的内方位元素（摄影机焦距 f、像主点在相片框标坐标系中的坐标 (x_0,y_0)）和外方位元素（投影中心在物方空间坐标系中的坐标值 x_s,y_s,z_s 和相片的三个空间姿态角）. 当内方位元素已知时，只用于测算外方位元素. 这些元素是摄影测量中的重要参数.

航摄相片纠正(aerial-photo rectification) 航摄相片处理的一个重要步骤. 即消除由于航摄相片倾斜角所引起的像点位移，限制地形起伏引起的投影差，并使影像归化至固定比例尺的技术. 成果为半色调的纠正相片或线划图. 它分为光学机械法、光学图解法和图解法. 现在主要采用光学机械法，使用的仪器称为纠正仪.

正射影像技术(orthophoto technique) 航摄相片处理的一种技术. 即制作各类地区正射相片图的技术. 有光学投影和电子投影两类. 前者用正射投影仪上的一个小缝隙与立体测图仪的测标相对应（缝隙下面放置感光材料晒像）. 又分联机和脱机作业两种方式. 联机作业时，缝隙与测标同步扫描移动，观测者在操作过程中只需不断改变测标的高程，使它始终与立体模型的表面相切，缝隙中的影像就随之改变比例尺，晒印得到目标物（如山地）的正射相片图；脱机作业时，在立体测图仪上观测时只记录了各扫描断面的高程数据，然后根据此记录，在电子计算机控制下，使缝隙作扫描晒像. 脱机方式灵活方便，得到普遍采用. 后者只能在自动化测图仪（毋需人眼观测的测图仪）上使用. 先将一张相片划分为若干小块（如相片上 9 mm×8 mm），用一个两次以上的数学模型来拟合. 对每一块小面积的立体模型，电子计算机计算出各个像素纠正后的位置，并控制飞点扫描器晒像，将中心投影的影像转变为正射相片图.

相片三角测量(phototriangulation) 航空摄影测量中加密测图控制点的一种方法. 是根据少数野外控制点，利用一条航带或某一摄区中各航摄相片之间内在的相关特性，用交会法确定测图所需控制点的平面位置和高程. 分为辐射三角测量和空中三角测量两种.

空中三角测量(aerial triangulation) 一种航空摄影测量方法. 即航空摄影测量中室内加密测图所需的高程和平面控制点的最常用方法. 例如，以像对建立单元的立体模型为基础，通过相邻模型重叠范围内连结点进行模型连结，构成一条航带或由几条航带构成的区域网，再根据少数野外控制点，进行三维坐标的相似变换（绝对定向）和非线性误差改正，即确定测图所需控制点的平面位置和高程. 该方法一般分为模拟空中三角测量和解析空中三角测量.

辐射三角测量(radial triangulation) 航空摄影测量中室内加密平面控制点的一种方法. 根据以相片的辐射中心为顶点在相片上量取的平面角与以辐射中心的地面相应点为顶点量测相应方向之间的水平角近似相等的几何特性，在像对的重叠范围内，可依据相片基线使同名辐射方向实行平面交会，依次连结，可构成航带的菱形三角锁或几条航带连结的辐射三角网，再根据少量地面控制点进行比例尺归化，从而确定各交会点的平面位置. 它适用于平坦地区的平面控制点加密.

解析空中三角测量(analytical aerial triangulation) 摄影测量中加密控制点的一种方法. 采用电子计算机进行计算，故亦称电算加密. 该方法能在一条摄影航带或若干相邻航带构成的区域内，根据少数野外控制点，确定测图所需控制点的平面位置和高程. 按平差单元的性质不同，有航带法、独立模型法和光线束法等三种常用的方法. 把若干条相邻航带连成整体进行平差的方法，称为区域网空中三角测量.

航带法区域网空中三角测量(strip block aerial triangulation) 解析空中三角测量的一种. 以一条航带网作为整体平差运算中基本单元的一种区域网加密方法. 作业过程是：

1. 各航带分别用连续法相对定向，构成自由比例尺的航带网.

2. 对第一条航带，根据足够数量的控制点进行

绝对定向.

3. 第二条航带则以本航带内的控制点和与第一条航带公用的连结点(其坐标值采用第一条航带绝对定向后的坐标值)为依据,进行第二条航带的绝对定向.

4. 其余航带用同样方法顺序地连成一个区域网,获得全区域统一系统的坐标值.

5. 按最小二乘法原理进行整体平差,其原则是:平差后控制点(作定向点用)坐标的平差值应等于它的大地坐标值,在相邻航带中分别求得公共点的坐标值应相等. 这种方法在整体平差计算中所确定的是各航带作非线性改正的待定参数(若用二次多项式作航带非线性变形改正,每航带有 15 个待定参数). 最后,计算各加密点的坐标值.

独立模型法区域网空中三角测量(independent model block aerial triangulation) 亦称独立模型法区域网加密. 解析空中三角测量的一种. 是以单元模型中各自形成的连结点和投影中心坐标作为一个整体,参加全区域平差计算的区域网加密方法. 单元模型可以由一个或两个甚至更多的像对构成. 先将各个单元模型按其前、后、左、右的连结关系,纳入统一的坐标系中,然后把这些坐标值作为观测值,按最小二乘法原理进行平差计算. 在整体平差过程中,使每个单元模型保持其内部形状的相对关系不变,而进行位移、旋转和缩放,共有七个参数. 平差的原则是:经整体平差后,作定向点用的控制点坐标平差值应等于它的大地坐标值,相邻模型间的两连结点在各个模型中求得的坐标平差值应相等. 整体平差过程就是确定各个单元模型的七个参数. 最后,计算所有加密点的坐标值.

光束法区域网空中三角测量(bundle block aerial triangulation) 亦称光束法区域网平差. 解析空中三角测量的一种. 利用在摄影时地面点、相应像点和摄影站点三点共线条件所建立的每条空间光线(一张相片为一个空间光线束)作为一个单元,进行全区域统一平差计算的区域网加密方法. 建立三点共线条件,需确定六个待定参数,此即每张相片作旋转和平移的六个外方位元素. 按最小二乘法原理进行区域网平差时,必须首先代入每张相片外方位元素的初始值和全区域各加密点坐标的初始值(采用较精确的初始值,可减少运算时间,提高计算成果的精度). 因为光束法列出的误差方程式中使用的观测值是量测的像点坐标 (x_i, y_i),所以该方法严格符合最小二乘法的要求,与其他加密方法相比,在理论上较严密.

光束法区域网平差(bundle block adjustment) 即“光束法区域网空中三角测量”.

联机空中三角测量(on-line aerial triangulation) 亦称在线空中三角测量. 解析空中三角测量的方法之一. 是一种在仪器上边观测、边计算实时地进行空中三角测量的作业方法. 以一台电子计算机与立体坐标量测仪相连结,作业员在立体坐标量测仪上观测立体像对,观测成果直接输入电子计算机,实时进行解析空中三角测量的各项计算. 优点是提高生产效率、及时发现和改正差错、可使用各类解析测图仪.

在线空中三角测量(on-line aerial triangulation) 即“联机空中三角测量”.

数字地面模型(digital terrain model) 简称 DTM. 地面点高程数据组成的一个集合. 测绘工作中,以密集的地形模型点的坐标 (x, y, z) 来表达地表面(或被测物体表面)形态的一种方式. 通常用规则的格网就可以记录被测点的行列序号和记录格网点的高程,应用比较方便,此时亦称数字高程模型. 格网点的坐标和高程主要用摄影测量方法测得,亦可由野外实测,并在其基础上,选用适当的函数式进行数据内插,加密至所需要的密度. 所获数据输入存储器(如磁带、磁盘等)中,借电子计算机控制数控绘图桌绘出等高线图,或控制正射投影仪制作正射影像相片图,通过一定的编码方式还可表达地物要素,是数字地图和地形测量数据库的重要依据. 在工程设计中可用来绘制断面图、计算工程量和进行各项经济指标的比较,为设计自动化开辟新的途径.

数字影像(digital image) 亦称数字相片. 影像的一种表示形式. 即以数字形式存在的影像. 航摄相片通过采样和量化过程,转化为大量的密集的灰度数字阵列,并记录和存储在磁介质中;或在航天摄影中采用电荷偶合器件(CCD)扫描得出,并记录和存储在磁带中的“影像”. 它存储方便、应用灵活,有利于航测自动化测图. 通过专门设备可将它转印成需要比例尺的相片. 还可直接输入数字化全自动测图系统,进行自动化测图,最后输出具有等高线的正射影像地形图.

数字相片(digital photograph) 即“数字影像”.

地形测量数据库(topographic data base) 一种测绘产品. 即地形测量信息系统的基本存储体. 存储的主要内容是数字地面模型和用代码形式存储地形图中所有的地物要素与有关数字和文字的注记. 数据库中的内容可随时适当地修改和补充,以正确地反映地面的现实状况. 根据用户需求可提供各种形式的地图或影像图. 主要用于存储国家基本图(比例尺为 1∶5000～1∶100000).

相片判读(photo-interpretation) 对相片中的物体进行分析识别的过程. 即根据航空或地面摄影相片上物体影像的形状、大小、色调、阴影、相关位置

和活动痕迹等特征，判释物体的性质. 在地形测量中，通过判读和实地调查，将判定的影像以地图的图式符号和必要的注记绘注在相片上. 相片判读在地质勘探、工程测量、土壤、环保、森林和军事等方面都有广泛的应用.

自动相片判读(automatical photo interpretation) 利用计算机进行相片判读的工作. 主要用电子计算机的专门软件，判释出与相片影像或扫描影像相应的地物的类别、特性、背景关系和它的动态变化等工作. 由于航空相片和从卫星上拍摄的相片，多数采用竖直摄影，所以地表面在相片上的构象存在着几何规律和相似性. 摄影时采用各种不同感光特性的摄影材料(如全色片、彩色片、红外片和假彩色片等)可以强调反映某部分光谱特性地物的特点. 还可采用多光谱摄影和雷达摄影. 自动判读处理时，参照在实地采样所取得的地物波谱记录以及典型调查的样片进行的对照、分析和判断来判读. 它在测绘、农业、林业、地质、资源调查以及军事侦察等各个领域中均有应用.

非地形摄影测量(non-topographic photogrammetry) 摄影测量的一种. 对非地形目标进行摄影并进行摄影测量处理，以确定其外形、状态和几何位置的技术. 内容包括静态物体的外形测量、动态目标的轨迹测量、燃烧爆炸与晶体生长等不可接触物体的实时测量. 广泛应用于建筑物、文物、医学以及变形测量等.

近景摄影测量(close-range photogrammetry) 非地形摄影测量的一种. 摄影距离从几米到 300m 范围内非地形目标的测量技术. 一般的特点是比例尺较大，量测精度要求较高. 可采用量测用摄影机，也可采用可调节主距的或普通的摄影机. 后者的内定向元素是未知的. 处理时，多数采用解析法或解析与模拟相结合的方法. 在建筑物测量、变形测量、工业测量、生物医学测量以及刑侦等方面应用甚多.

双介质摄影测量(two-medium photogrammetry) 一种摄影测量方法. 被摄物体与摄影机处于不同介质中的摄影测量技术. 双介质摄影的成像光线必定穿过两个不同的介质(如空气和水)，这就使摄影测量处理时需要考虑成像光线在介质分界面处的折光. 当摄影机置于空中向水内摄影时，水面就是两介质的分界面；若摄影机在水下摄影时，则摄影机物镜的主平面是两介质的分界面. 用于测绘海底地物和研究水中物体.

直接线性变换(direct linear transformation) 摄影测量中进行相片数据处理的一种方法. 一般用于各类非量测用摄影机(如普通照相机、高速摄影机等)拍摄的未知内方位元素的相片. 其实质在于建立坐标仪上量测的像点坐标和物方(控制点和其他量测点)空间坐标的直接变换关系式. 惯用的内、外方位元素等都隐含在这一关系式中的若干个待定参数里，因此可跳过求解内、外方位元素的过程，计算时毋须应用未知数的起始近似值.

数字化测图(digital mapping) 摄影测量中的新测图方法之一. 可将数字地面模型、地图要素或影像以数字方式记存于磁带中，经电子计算机处理，输出线划图或正射影像地图. 按数据获取的方式分为两类：

1. 用立体测图仪或解析测图仪得出数字地面模型，并输入地物符号与注记等的信息编码，或用数字化仪器根据已有地形图获得上述数据和信息.

2. 用数字相片或将相片按影像灰度等级进行数字化，再在立体像对中用电子计算机相应软件进行相关运算和摄影测量的计算，取得数字地面模型和正射影像地图.

航测自动化测图(automatic mapping in aerial photogrammetry) 一种航摄测图新技术. 即摄影测量中不用人眼观测而进行自动测图的技术. 关键是把影像相关技术用在立体测图中自动寻找同名像点. 影像相关的方法主要有电子相关和数字相关两种，其原理都是以相关函数为基础的. 此外，还用一系列透镜组成的相关器进行光学相关. 影像相关的设备是以一台电子计算机为主体，连结许多有关部件构成，故一般不称为仪器，而称为全自动化测图系统.

全自动化测图系统(fully automatic mapping system) 见“航测自动化测图”.

单元立体模型的相似变换(similarity transformation of unit stereo-models) 一种相片变换方法. 即连续的多张相片数据的整体变换方法. 在欧氏空间内，一个或多个相邻立体像对构成的立体模型在三维空间的线性变换公式，由于变换前后保持立体模型的相似性，故称为三维相似变换. 其基本关系式为

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \lambda \cdot R \cdot \begin{bmatrix} \overline{X} \\ \overline{Y} \\ \overline{Z} \end{bmatrix} + \begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix},$$

式中$(\overline{X},\overline{Y},\overline{Z})$是单元模型中各连结点或投影中心点的重心化摄影测量坐标；(X,Y,Z)为相似变换后点在指定的地面三维直角坐标系中的坐标；(X_G, Y_G, Z_G)为单元模型坐标平移值，即单元模型重心点的地面坐标；λ 为单元模型的缩放值；R 为单元模型的旋转矩阵，反映两个直角坐标系的基变换. 该模型有三个独立未知数.

单元立体模型的相似变换用于模型连结、绝对定向和独立模型法区域网平差，是解析摄影测量的

基本公式之一.

摄影测量内插(photogrammetric interpolation) 摄影测量数据处理的一种方法.是根据一系列参考点(或称数据点)上的某些信息拟合一曲线或曲面,从中求出其他点上的相应信息的数字处理过程.

摄影测量内插方法分为整体函数法、局部函数法和逐点内插法.整体函数法多采用 n 次多项式函数,与参考点数有关.局部函数法有线性内插、局部多项式函数、Akima 内插、双线性内插、样条函数、有限元内插等方法.采用局部函数法时,需要考虑是否顾及相邻分块函数间的连续性问题.逐点内插法是对每一个待定点定义一个新的内插函数,可用加权平均值法、移动拟合法和最小二乘配置法.

摄影测量内插在解析摄影测量中有多方面的应用.曲线的拟合与内插用于绘制等值线图和数字测图中各类光滑曲线;曲面的拟合与内插用于系统误差改正、数字地面模型的内插等.

影像采样(image sampling) 一种影像数字化的过程和技术.是按一定间隔对影像灰度进行采样,获得影像灰度矩阵的过程.可以对影像平面在空间域内和在频率域内用卷积和乘法的过程来确定合适的采样间隔.根据仙农(Shannon,C. E.)采样定理,采样间隔 Δx 应满足条件(以一维为例)

$$\Delta x \leqslant \frac{1}{2W},$$

式中 W 为频率限值.就是说,当采样间隔能使函数 $f(x)$ 中存在的最高频率中每周期取两个样本时,则可根据采样数据完全地恢复原函数 $f(x)$,故称 W 为截止频率或奈奎斯特频率.考虑到原始影像有噪音,采样光点也不可能是理想的光点,所以实际采样间隔应保证在最高频率中每周期至少取到三个样本.

影像采样的仪器称为影像数字化仪或影像扫描仪.采样间距可以调节,通常又称为像元.像元大小除了遵循采样定理外,还要考虑到影像处理的要求与条件.

影像重采样(image resampling) 一种影像数据处理方法.即影像数据重新组织过程中的灰度处理方法.影像采样是按一定间隔采集影像灰度数值的,当欲知不位于采样点上的原始函数的数值时,就需要利用已采样点进行内插,称为重采样.在数字摄影测量和遥感中,实际的重采样发生在影像旋转、核线排队、数字纠正和多重影像的复合过程中.一般采用的重采样方法有双线性插值法、双三次卷积法和最邻近像元法.前两种方法几何精度较高,后一种方法最简单快速,且不破坏原始影像的灰度信息,只是几何精度只能达到±0.5个像元.

数字影像匹配(digital image matching) 一种影像处理方法.即在包含有相同物体目标的两幅数字影像中搜寻共同目标的计算机处理方法.是利用电子计算机通过数字图象处理的方法辨认不同数字影像上的同名目标(点、线、面等)的自动处理技术.如果对有重叠的两张数字影像进行匹配,通常又称为数字影像相关,它将代替人工实现自动立体观测,用于自动空中三角测量、自动生成数字地面模型和制作数字正射影像.数字影像匹配也可以在影像目标与其原始的参考目标之间进行,如自动识别相片上的框标、地图上的公里格网等.用数字影像匹配方法寻找不同时相、不同传感器遥感影像之间的同名目标,通常又称为数字影像配准.

按照图象处理从低到高的水平进行数字影像匹配,可分为基于灰度的影像匹配、基于特征的影像匹配和基于解译的影像匹配.高水平的影像匹配需要借助于一定的知识进行图象理解,在此基础上实现影像匹配.三个层次的方法可以组合使用.计算机视觉中使用的模板匹配可归属于基于特征的匹配,而所谓关系匹配则可视为高水平的基于解译的匹配方法.

数字影像匹配是摄影测量与遥感以及计算机视觉中自动重建目标几何模型的重要手段和关键技术,是当今一个重要研究课题.

数字影像相关(digital image correlation) 见"数字影像匹配".

数字影像配准(digital image registration) 见"数字影像匹配".

基于灰度的影像匹配(grey value based image matching) 一种数字影像匹配方法.根据左片目标区和右片搜索区中的影像灰度信息寻找最相似的同名目标,称为基于灰度的影像匹配.因为它总是在一个窗口内进行,又是一种基于面积的影像匹配,以示与基于特征的方法相区别.

基于灰度的影像匹配方法有利用相关系数的核线相关方法、阿克曼(Ackermann,F.)提出的最小二乘高精度数字相关法,还有顾及共线方程的最小二乘相关法、铅垂线轨迹法、多点动态规划算法和有限元多点最小二乘匹配法等,目前多采用多条核线、多点带制约条件的整体动态规划松弛算法.

美国的海拉瓦(Helava,U.)等人,提出了直接将影像灰度与其地面目标的灰度进行匹配的物方空间影像匹配方法.在物方空间影像匹配中,除了灰度方程式外,还有几何方程式,所求的未知数中还包括地元的数字正射影像灰度值.

基于灰度的影像匹配可以达到很高的精度,但它没有充分利用影像所包含的全部信息,对初始值要求高,而且需要解决质量控制和成果自动诊断等

问题.

基于面积的影像匹配(area based image matching)　即"基于灰度的影像匹配".

基于特征的影像匹配(feature based image matching)　一种数字影像匹配方法. 是在两张相片上利用影像分析方法提取点、线、面特征,再找出两相片间相匹配的同名特征,从而实现自动立体量测的一种方法.

在数字摄影测量中,常用的特征提取的有效算子有 Moravac 算子、Dreschler 算子和 Förstner 算子. Moravac 算子的出发点是基于理想的特征点应在所有方向上有很大的反差,而理想的特征线,则仅在其垂直线方向上才有很大的反差. Dreschler 算子则利用密度函数的主曲率、高斯曲率和主方向夹角来提取特征. Förstner 算子是从特征点所在影像窗口几何变换协方差阵中求出长短半轴比和权值,选那些长半轴差别尽可能小,而权尽可能大的点为特征点. 左右片特征点之间的匹配一般采用稳健估计方法,以自动剔除特征提取中的差错.

基于特征的影像匹配至少有两个好处:其一是它不需要很精确的初始值,可以在较大范围内寻找特征;其二是可以用一些快速算法,方便快速地实现,而且出错和失去匹配的可能性较小. 缺点是影像匹配的精度只达到±1 个像元. 因此,它通常作为粗一级的影像匹配,用来为基于灰度的影像精匹配提供初始值.

针对工业摄影测量中目标结构特征丰富的特点,人们从基于特征的影像匹配出发,提出和发展了特征摄影测量.

自检校平差(self-calibrating adjustment)　一种摄影测量平差方法. 即能自动消除测量中系统误差的平差. 是利用引入附加的未知参数,在测量平差中自动补偿观测值中所包含的系统误差的一种方法. 由于它针对可能的系统误差源假设了一个系统误差模型,在平差中求解其他未知数的同时也求解出该系统误差模型中的参数,因此它等效于在实验室或用试验场对系统误差的检校,故取名自检校平差.

在摄影测量点位测定中引入自检校平差的原因在于相片系统误差不可能事先完全精确求出,它乃是成像系统、大气条件、飞行条件及作业环境的综合影响结果. 比较著名的系统误差模型有模拟物理因素作用的 21 个参数的布朗模型、基于正交多项式的 12 个参数的埃伯勒模型和 44 个参数的格林模型,以及由康尼斯内(Konecny,G.)等人提出的包含傅里叶系数的附加参数模型. 李德仁的研究指出,自检校平差是从有噪声(偶然误差)的观测值中提取信号(系统误差),其效果主要取决于信噪比.

为了附加参数的选择,可以在自检校平差中进行统计检验,主要包括显著性检验、可测定性(内部可靠性)检验、灵敏度(外部可靠性)检验和相关性(可区分性)检验. 当附加参数选择得过多,会引起方程组状态急剧变坏,解的精度严重下降,称为过度参数化. 改善平差的几何强度,将附加参数处理成带权观测值或采用岭估计方法求解均可克服过度参数化.

自检校平差用于高精度摄影测量或大地测量点位测定中,是补偿和消除系统误差影响的一种有效方法.

粗差检测(gross error detection)　一种误差处理理论和方法. 即发现和排除大误差观测值的理论和方法. 在测量平差过程中,自动地发现和剔除观测值中包含的粗差,即异常观测值. 由于粗差是一种模型误差,它可以从两个角度去理解,从而引出两类粗差检测方法.

第一,含粗差的观测值 l_i 可以视为与其他同类观测值 l_j 具有相同的方差,但期望不同,即期望异常(如图 1 所示).

$$\begin{cases} l_i \sim N(E(l) + \nabla \tilde{l}_i,\ \sigma^2), \\ l_j \sim N(E(l), \sigma^2) \quad (j \neq i) \cdot \end{cases}$$

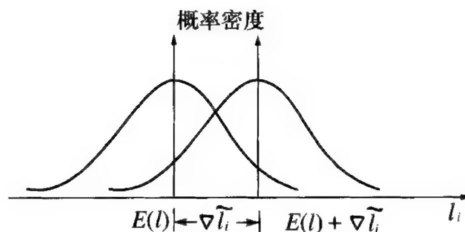


图 1

由此出发,利用假设检验理论,可以构造一维或多维的统计检验量来检测粗差. 其中,已知单位权方差下的单个粗差检测方法就是熟知的由巴尔达(Baarda,W.)提出的数据探测法.

第二,含粗差的观测值 l_i 可以看做与其他同类观测值具有相同期望,但方差不同,即方差异常或称方差膨胀(如图 2 所示).

$$\begin{cases} l_i \sim N(E(l),\ \sigma_{l_i}^2) \quad (\sigma_{l_i}^2 \gg \sigma^2), \\ l_j \sim N(E(l),\ \sigma^2) \quad (j \neq i), \end{cases}$$

由此出发,根据观测值的权应该与观测值的方差成反比的原理,可以利用每次平差的残差结果,按一定的权函数公式,计算下一次迭代中各观测值的权. 当权函数选择适当时,含粗差观测值的权将愈来愈小,从而达到剔除粗差的目的. 这类方法称为选权迭代法.

对于具有大量观测数据的近代平差系统,在平差过程中进行粗差检测是十分重要的,也是可能的.

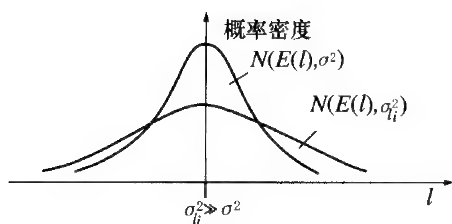


图 2

数据探测法(data snooping)　一种粗差检测方法. 以期望异常为备选假设,对原平差模型进行统计检验以发现观测值粗差的一种方法. 最先提出数据探测法的是荷兰学者巴尔达(Baarda,W.). 在观测值不相关的情况(权矩阵为对角线矩阵)下,对于已知单位权方差因子,他导出了标准化残差

$$w_i=\frac{|v_i|}{\sigma_0\sqrt{q_{v_{ii}}}}=\frac{|v_i|}{\sigma_{v_i}}=\frac{|v_i|}{\sigma_{l_i}\sqrt{r_i}}$$

作为正态变量的检验量. 式中 w_i 为标准化残差,v_i 是第 i 个观测值改正数,$q_{v_{ii}}$和 σ_{v_i}分别是 v_i 的权倒数和中误差,σ_{l_i}是 l_i 的中误差,r_i 是第 i 个观测者的多余观测分量. 当标准化残差大于检验的临界值时,则认为该观测值可能含有粗差. 随后,许多学者又将数据探测法扩充,提出其他方法来检测粗差. 中国李德仁和於宗俦等导出了一般情况下单个或多个粗差的统计检验量.

选权迭代法(iteration method with variable weights)　消除测量中观测值粗差的一种方法. 即将含粗差的观测值视为随机模型中方差异常大的子样的一种剔除粗差的迭代平差方法. 由于粗差未知,平差仍从惯常的最小二乘法开始. 但是,在每步平差之后,根据残差向量和其他有关参数,按所选择的权函数,计算每个观测值在下一次迭代平差中的权. 如果权函数选择得当,且在平差几何下粗差可以定位,则含粗差观测值的权将愈来愈小,并趋近于 0. 迭代终止时,相应观测值的残差将直接代表粗差的大小,而平差计算结果将不受粗差的影响. 这样便实现了粗差的自动定位和改正.

在选权迭代法中,权函数的选择极为重要. 为此,各国学者提出了许多不同的权函数形式. 按其内容可区分为权是残差 v_i 的函数、权是标准化残差 w_i 的函数或权是观测值方差估值的函数. 按其形式分,常用幂函数或指数函数形式. 一些主要的权函数选择方法有最小范数迭代法、汉本法、丹麦法、周江文法以及李德仁法等.

选权迭代法可以较为方便地加到各种已有的平差程序中,而且可以同时剔除多个观测值粗差,在实际中应用较为广泛.

可靠性理论(reliability theory)　研究测量数据处理中模型误差的理论. 测量学中的模型误差指的是所建立的模型与客观存在的事物之间的差异. 它包括偶然误差、系统误差和粗差三类不同的误差. 处理偶然误差的方法是数学家高斯(Gauss,C. F.)于 1809 年提出的最小二乘法,之后一直没有完善的处理系统误差和粗差的理论,结果是:如果粗差或系统误差不能正确地被发现、被改正或被消除,势必影响最小二乘法平差的结果,使之变得不可靠而达不到人们所预期的理论精度. 因此,可靠性研究变得十分重要.

可靠性研究有两大主要任务. 一是从理论上研究平差系统发现和区分不同模型误差的能力以及不可发现、不可区分的模型误差对平差结果的影响. 另一是从实际中寻求在平差过程中自动发现和区分模型误差并确定其所在位置的方法.

可靠性研究的理论是建立在数理统计的假设检验的基础上. 经典的假设检验理论是 1933 年由奈曼(Neyman,J.)和皮尔逊(Pearson,K.)提出的. 在测绘界,可靠性理论是由荷兰巴尔达(Baarda,W.)在 1967—1968 年建立的. 该理论从单个一维备选假设出发,研究平差系统发现单个模型误差(粗差、系统误差等)的能力和不可发现的模型误差对平差结果的影响. 前者称为内部可靠性,后者称为外部可靠性. 1980 年前后,德国学者弗斯特勒尔(Förstner,W.)和科克(Koch,K. R.)将巴尔达的可靠性理论推广至单个多维备选假设的情况. 1983 年,弗斯特勒尔第一次提出模型误差的可区分性问题,并从两个一维备选假设出发,导出了可区分性本质上取决于两个检验量之间的相关系数的结论. 中国李德仁从高斯-马尔可夫模型所含的两个多维备选假设出发,于 1985 年发展了可靠性理论,提出了可区分性理论,称为扩展的可靠性理论. 该理论可研究系统发现并区分不同模型误差的能力,以及不可能与其他模型误差相区分的模型误差对平差结果的影响,从而为系统误差的区分、粗差的定位以及形变与粗差或系统误差的区分提供了基本理论和定量的尺度.

测量平差中的可靠性理论可以进一步发展,用于计算机视觉、数字图象处理、影像理解和专家系统中,最近形成的稳健计算机视觉就是例证.

扩展的可靠性理论(extended reliability theory)　见"可靠性理论".

空间信息系统(spatial information system)　一种信息处理与分析决策系统. 即以空间分布为特征的计算机信息系统. 数据库是以某种基本的数据集合的存储为中心,通过管理程序科学地组织和管理这些数据,并借助各种应用程序来完成各种输入和输出任务的计算机系统. 信息系统是以某种或某些数据库为基础,加上必要的采集、分析、应用、显示和输出的硬软件,以及人机交互界面组成的一个能回答有关某个或某些领域内任意问题,并做出科学

决策的计算机系统.空间信息系统则是基于空间分布数据的信息系统.最主要的空间信息系统是地理信息系统(GIS).其他还有土地信息系统(LIS)、医疗空间信息系统及数字表面模型(DOM)构成的工业用信息系统等.从计算机角度看,空间信息系统由计算机硬件、软件和空间型数据库组成.从系统结构和流程看,它包括空间数据采集、存储、加工、处理、表达和输出等相互联系的部分.

实时摄影测量(real-time photogrammetry) 一种新的摄影测量技术.即基于数字摄影的摄影测量数据实时处理技术.通常的摄影测量包括摄影、处理、量测或测图三个过程.因此,它尽管可摄得瞬时的影像,却不可能瞬时或实时地提供摄影测量结果.这对于非地形测量应用,如工业自动量测和控制过程、视觉机器人系统、动态变化监测过程或实时医疗诊断系统等是很不方便的,而且摄影测量需要训练有素的作业员进行作业,因此不能实现自动化.

随着电子计算机的发展,摄影测量的信息处理更多地采用数字方法,尤其是高分辨率的CCD(电荷耦合器件)数字摄影机的问世,使得数字影像可直接实时地输入电子计算机,从而免去了摄影处理的过程,为实时处理创造了必要的条件.所谓实时摄影测量,就是在人们允许或要求的极短时间内实现影像信息的获取、处理、加工和表达,并可直接与各种机助设计(CAD)和机助操作(CAM)系统相连结或对物体进行实时反馈操作或控制的摄影测量过程.

实时摄影测量系统至少包括两个自动而实时的作业过程,即实时影像数据采集和实时摄影测量处理,这都依赖于现代的数字摄影和摄影测量技术,而且还要做到快速而实时.

实时摄影测量是一门正在积极研究和开发的学科,有极其重要的应用前景.

遥感(remote sensing) 一般指一种现代探测地球表面信息的技术体系.广义而言,泛指各种非接触的、远距离的探测技术.狭义而言,主要指从远距离、高空,以至外层空间的平台上,利用可见光、红外、微波等电磁波探测仪器,通过摄影或扫描、信息感应、传输和处理,从而识别地面物体的性质和运动状态的现代化技术系统.任何物体在绝对温度零度(−273℃)以上都会反射或辐射不同波长的电磁波.人的眼睛或普通照相机,只能感受其中的可见光谱段,但特殊的遥感仪器却能把紫外、红外或微波的信息强弱及其空间差异记录下来,经过电子计算机和光电设备加工处理,再现这些物体的影像,变成人眼可以识别的图形,甚至按照专家系统的分析,直接输出结论性的专题地图.1959年,由人造卫星上发回了第一张地球相片.1960年,从“泰罗斯”与“雨云”气象卫星上,获得了全球的云图.1962年,美国密执安大学召开专题讨论会,会议取名“环境遥感”,公布了侧视雷达和红外扫描图象.同时,美国宇航局喷气推进实验室开始了对单色数字图象处理的研究,珀杜大学发展了彩色和多谱段图象处理技术.从此,“遥感”就成为从高空探测地球表面及其环境的信息获取、处理及其应用技术的专门术语.遥感在20世纪30年代航空摄影和判读的基础上,随着太空技术、电子计算机和地球科学的发展,产生了质的飞跃.它的主要特点有:

1.从以飞机为主要运载工具的航空遥感发展到以人造卫星为主要运载工具的太空遥感,人们开始从一个新的高度——太空,来观测地球.

2.超越人眼所能感受的可见光的限制,延伸了人的感官.

3.快速、及时地反映现象,用来对比分析环境的动态变化,从而赢得预测、预报的时间.

4.广泛吸收激光、光纤、全息等技术成就,涉及天文、地学、生物学等科学领域.

概括地说,遥感是运用物理手段、数学方法和地学规律的现代化综合性探测技术,它为经济建设、资源勘测、环境监测、军事侦察提供了现代化的技术手段,反映一个国家太空科学技术的进展、计算机技术的水平、地学科学的理论储备以及对资源、环境科学管理与预测、预报的能力.同时,也将是高技术开发和信息时代的新兴行业.

地面遥感(ground remote sensing) 一种基础性遥感技术.即在地面进行的地物波谱量测.主要指利用近地面的高塔、长臂车或地物波谱测试仪器对不同环境下各种地物的反射、发射波谱或多种地物的综合波谱进行量测,获取地物波谱特性的信息;并开展研究,以便从遥感影像中识别各类地物,为选择合适波段的传感器提供基础数据,进行遥感影像辐射校正等.因此,地面遥感通常是整个遥感系统的基础研究.

航空遥感(airborne remote sensing) 遥感技术的一种类型.泛指从飞机、飞艇、气球等空中平台对地观测的遥感技术系统.包括利用机载的各种遥感器:航空摄影仪(相机)、电视摄相机、扫描仪、散射/辐射计、雷达等,涉及紫外、可见光、红外、微波等多谱段.有些是多种遥感器组合在一起的实验型遥感飞机,也有一些搭载其他探测仪器专用于海洋、森林、探矿等业务型遥感飞机.除各种机载遥感仪器外,还搭载航空磁力仪、航空重力仪等.按飞机航行高度,分为低空(600～3000m)、中空(3000～10000m)和高空(10 000m以上)三级.此外,还有超高空侦察机(U-2)和超低空的遥控飞机都可以用于航空遥感,获取不同空间分辨率的相片或图象.1858年,法国人在气球上获得第一幅巴黎街区的相

片.后来气球让位于飞机,摄影镜头和胶片的改进也很快,1935年,开始有彩色摄影,1943年设计了红外彩色胶片.由于航空相片的比例尺较大,分辨率高,可以进行立体量测,即使在太空遥感高度发达的国家,航空遥感仍然是不可缺少的组成部分.航空遥感应用于大比例尺系列制图、地理信息系统建库和太空遥感的模拟试验和仪器检测,应用于城乡土地资源、森林、草场资源和区域矿产资源的预测和详查,应用于工程规划设计等.这些都已纳入作业规程,形成生产行业.一般航空遥感与单一的地面作业相比,可节约经费、加快速度、减轻劳动强度、克服自然环境的障碍.中国于20世纪30年代引进航空摄影,50年代开始应用于林业、水利、地质部门,60年代建立了航空相片综合利用研究室,70年代完成了覆盖全国的航空摄影测量与制图,积累了多年重复摄影的资料,为研究河道、海岸、湖泊、城市环境变迁提供了重要的资料.

航天遥感(astronautic remote sensing)　亦称太空遥感.遥感技术的一种类型.与卫星遥感大同小异.泛指利用各种太空飞行器为平台的遥感技术系统.以地球轨道人造卫星为主体,包括载人飞船、航天飞机和太空站.有时也把各种行星探测器包括在内,涉及空间天文学的一部分内容,但不包括平流层以下的火箭、气球、飞艇和超高空飞机.其主要特点是:由运载火箭发射太空平台,一般超过第一宇宙速度,按一定的轨道高速运行.采集遥感信息有四种基本形式:

1.宇航员操作,如在"阿波罗"飞船上利用组合相机拍摄地球相片.

2.卫星舱体回收,如中国科学技术卫星和苏联"联盟"号拍摄的卫星相片.

3.通过扫描将图象转换为数字编码,传输到地面接收站,记录于高密度磁带并回放为图象.

4.卫星数据采集系统收集地球或其他行星、卫星上定位观测站发送的观测信号,中继传输到地面接收站.

航天遥感技术的发展,使人类观察地球达到了新的高度.极地轨道卫星的高度一般约1000km,"陆地卫星"的一幅多谱段图象,覆盖地面面积34000km²,相当于海南岛或台湾岛的面积.静止气象卫星轨道高度约36000km,可覆盖到南北纬40°以内,东西经距70°左右.通过航天遥感,人们获取了许多前所未有的客观图象,诸如大气环流、海流、板块构造缝合带、大裂谷等,已成为太空技术和平利用的重要组成部分.

可见光遥感(visual remote sensing)　遥感技术的一种类型.主要利用人类眼睛可见的谱段(波长0.4～0.7μm)的遥感技术的统称.一般采用感光胶片或光电探测器作为感测元件,以摄影方式为主体.利用前后或左右重叠的立体摄影像对,通过立体量测与制图仪,获取地面的高程数据和高精度的空间定位数据,满足地形图测绘的要求.中国为科学技术实验卫星研制了全景摄影仪,回收的相片星下点附近的地面分辨率达到10m左右;法国资源卫星(SPOT)全色扫描影像地面分辨率为10m,并可利用轨道侧向重叠进行立体量测;苏联"联盟"号飞船影像的地面分辨率达到5m;美国的低轨侦察卫星据说可达到0.3m的分辨率.可见光遥感应用广泛,特别是航空摄影积累了宝贵的历史资料,覆盖面积约达陆地面积的70%,对环境动态变化的对比分析,具有重要价值.

红外遥感(infrared remote sensing)　遥感技术的一种类型.利用红外谱段的遥感技术.通过红外敏感元件,控测物体的红外辐射能量,显示目标的辐射温度或热场图象的遥感技术的统称.主要仪器有单通道红外辐射计、红外热像仪、多通道红外扫描仪等.在多谱段扫描仪中,除可见光谱段外,往往也包括1～2个或几个红外谱段的通道.红外遥感按波长分为三个谱段:波长0.76～3.0μm称为近红外,可用感光胶片直接感测;波长3.0～6.0μm称为中红外;波长6.0～15μm称为远红外.中、远红外遥感具有昼夜工作的能力,对于夜间侦察、叶绿素含量与生物量估算、水体热污染、地热与地下水勘探,具有特别重要的意义.中国研制的红外遥感仪器,不仅广泛运用于航空遥感,而且满足了气象卫星的需要,在林火、海冰、渔场、作物长势等方面获取了重要的信息.

微波遥感(microwave remote sensing)　遥感技术的一种类型.利用波长1～1000mm电磁波的遥感的统称.通过接收地面物体发射的微波辐射能量,或接收遥感仪器本身发出的电磁波束的回波信号,对物体进行探测、鉴别.前者称为被动(无源)遥感,如微波辐射计;后者称为主动(有源)遥感,如雷达、散射计、高度计等,这些仪器本身发出电磁波,所获得的回波与日照的变化无关.微波遥感的特点是:对云层、地表植被、松散沙层和干燥冰雪具有一定的穿透能力,又能夜以继日地全天候工作,不仅在热带雨林、两极冰盖和干旱沙漠地区发挥了不可代替的作用,而且在海洋探测方面也占重要地位.国际上已有六种航空侧视雷达投产.一种星载的聚焦合成孔径侧视雷达,预计在690km的轨道高度上,地面分辨率可达4m.微波遥感应用潜力很大,技术发展方兴未艾.未来的卫星遥感计划中已将微波遥感列为重要内容,如1995年加拿大发射RADARSAT雷达卫星,美国在2000年2月"奋进号"航天飞机SRTM计划中,用双天线SAR(合成孔径雷达),在11天内获取了地球陆地北纬60°到南纬56°间80%

地区的立体雷达数据.

无源遥感(remote sensing without man-made radiation source)　见“微波遥感”.

有源遥感(remote sensing with man-made radiation source)　见“微波遥感”.

多谱段遥感(multispectral remote sensing)　遥感技术的一种类型. 即同时利用多个可见光和红外谱段的遥感技术,包括多谱段摄影、多谱段扫描和细分光谱仪成像. 多谱段摄影机或扫描仪无论是装备在遥感飞机上,还是装备在人造地球卫星上,都能获得光谱分辨率较高、信息量较丰富的图象和数据. 多谱段摄影机上装有 4～9 个镜头,分别摄取同一目标的不同谱段的光谱,记录在同一底片或各自的底片上;然后根据需要,选择其中两个或三个谱段的底片,组合曝光成一张新的相片. 按自然光谱合成的,称为天然色相片;人工任意调配颜色的,称为假彩色合成相片. 对摄影或扫描图象的色度、密度进行计算机处理、数值分类、分级与分析,可以得到不同的结果,从中提取土地覆盖与土地利用、土壤与地质、水体污染与城市建设的有效信息,为资源勘探和环境监测服务. 美国“陆地卫星”上的多谱段扫描仪选用 2 个可见光谱段和 2 个近红外谱段;专题成像仪选用 3 个可见光谱段和 4 个红外谱段. 法国“斯波特”卫星选用 2 个可见光谱段和一个红外谱段.

天然色相片(natural color photograph)　见“多谱段遥感”.

假彩色合成相片(pseudo color composite photograph)　见“多谱段遥感”.

空间分辨率(spatial resolution)　亦称地面分辨率. 一种遥感技术性能指标. 即是遥感仪器所能分辨的最小目标的大小. 通常在光机扫描成像仪中,探测器所接收的辐射通量是一个瞬时视场发射到成像仪的总的出射辐射通量,而不管这个瞬时视场内有多少个不同性质的目标,也就是说,遥感器不可能分辨出小于瞬时视场的目标. 因此,光机扫描遥感器的空间分辨率就是它的瞬时视场. 通常通过增大光学系统焦距或减小探测元件的尺寸来提高遥感器的空间分辨率,而光学遥感仪器的最大空间分辨率又受光学系统的衍射极限限制.

地面分辨率(ground resolution)　即“空间分辨率”.

光谱分辨率(spectral resolution)　一种遥感技术性能指标. 即机载和星载遥感器的一项性能指标. 指遥感器在接收目标辐射的光谱时,能分辨的最小的波长间隔;或是指对两个不同辐射源的光线波长的分辨能力. 能分辨的波长间隔越窄,则光谱分辨率就越高.

遥感器中分光元件的角色散率、会聚镜的焦距及探测器阵列相邻探测元的间距都决定系统的光谱分辨率. 光谱分辨率愈高,愈能获得物光的精细光谱特征,但元件接收的能量愈小.

电磁波(electromagnetic wave)　一个基本物理概念. 即在空间传播的交变电磁场. 这些电场和磁场以光速传播并且彼此垂直,同时还与传播方向垂直. 无线电波、微波、红外线、可见光、紫外线、X 射线、γ 射线都是电磁波,不过它们产生的方式不同,波长也不同. 电磁波有时特指天线发射或接收的无线电波.

电磁波谱(electromagnetic spectrum)　一个物理概念. 即电磁辐射按波长或频率数值大小顺序排列的电磁波序列. 它从最长的无线电波直到最短的 γ 射线.

可见光(visible light)　一个特殊光谱段. 即能引起人眼视觉的电磁波. 波长范围为 380～780ns. 落在可见光谱段的太阳辐射能量约占总辐射能量的一半,同时,大气对此谱段的透过率也极高,是极其理想的遥感谱段,也是人类用得最多、了解得最清楚的谱段.

近红外(near infrared)　一个特殊光谱段. 即电磁波波谱中与可见光段的红端最邻近的谱段. 遥感中,通常将 0.76～3.00 μm 波长范围划为近红外谱段,简称近红外. 遥感技术中又称该谱段为反射红外. “陆地卫星”多谱段扫描仪的两个近红外谱段为 0.7～0.8μm 及 0.8～1.1μm. 植被在该二谱段具有强反射的特点. “陆地卫星”4 号和 5 号的专题成像仪,具有三个近红外谱段,即 0.76～0.90μm,用于水体成像、探测水中生物含量及岩石、土壤中的含铁情况;1.55～1.75μm,用于探测植物含水量及土壤湿度,区分云与雪等;2.08～2.35μm,用于探测地质岩石类型,研究粘土类矿物、碳酸盐类矿物以及区域地质制图和大型热液蚀变带的识别研究等.

反射红外(photographic infrared)　即“近红外”.

中红外(middle infrared)　一个特殊光谱段. 即指遥感中 3～6 μm 谱段的电磁波. 遥感技术中,为了探测高温辐射源发出的电磁辐射,如监测火山活动、森林火灾等,常选用 3.0～5.5 μm 谱段(大气窗口之一)作为遥感器的探测通道.

远红外(far infrared)　一个特殊光谱段. 即指遥感中 6～15 μm 谱段的电磁波. 其中 8～14 μm 波段是大气窗口之一,亦是遥感技术中用于探测地球表面物质自身辐射的主要谱段,是热红外扫描仪的主要探测通道.

热红外(thermal infrared)　一个特殊光谱段. 即指遥感中 3～14 μm 谱段的电磁波. 地物在常温(约 300K)下热辐射的绝大部分能量位于此谱段.

在此谱段，地物的热辐射能量大于对太阳的反射能量，遥感器接收到的信息主要来自地物的热辐射.

微波(microwave) 一个特殊电磁波谱段. 即一般指波长为 1mm 到 1m 波段的无线电波. 微波具有极强的穿透云雾的能力，所以微波遥感器具有全天候、全天日工作能力. 微波对某些介质，如干沙、冰等有一定的穿透能力，能在一定程度上反映地表下的信息. 一些在可见光、红外波段不易区分的地物在微波波段可明显地区分. 微波波段的遥感器在军事、海洋、资源探测等各领域均有广泛应用. 此外，还用于卫星通信、电视、射电天文、雷达等方面.

散射(scattering) 一种电磁波传播现象. 即电磁辐射在非均匀媒质或各向异性媒质中传播时改变原来传播方向的现象. 散射后电磁辐射传播方向与原入射方向小于 90°的，称后向散射；90°至 180°之间的，称前向散射. 引起散射的粒子尺度远小于入射电磁波波长的，称瑞利散射，其特点是散射强度与波长的四次方成正比，其前向散射与后向散射的强度相同. 大气中分子运动的涨落所引起的散射即属瑞利散射. 如果粒子的尺度与入射波长相近，则称米氏散射，其主要特点是前向散射远大于后向散射，霾的散射即是. 如果粒子尺度远大于入射波长，则称无选择性散射，其散射强度是各向同性的，强度与波长无关，如云层中水滴的散射.

散射对遥感数据传输的影响极大，如减小太阳光直射的强度，引起漫入射的天空光，使遥感器接收到的地面辐射强度减小，增加大气散射引起的天空辐射，使遥感影像的反差降低. 在大气窗口内，散射是引起信息畸变的主要原因. 在雷达遥感中，对物体散射特性的研究是判读雷达图象的基础.

后向散射(backward scattering) 见“散射”.

前向散射(forward scattering) 见“散射”.

米氏散射(Mie scattering) 见“散射”.

无选择性散射(non-selective scattering) 见“散射”.

吸收率(absorptivity or absorptance) 一个电磁波辐射的物理量. 即物质吸收的辐射通量与总入射辐射通量之比值. 常用 α 表示，是个量纲一的量.

透射率(transmissivity or transmittance) 一个电磁波传播的物理量. 即透过物体的电磁波强度与总入射波强度之比. 对于同一物体，透射率是波长的函数. 例如，可见光对清水的透射率，随波长增加而迅速减小. 常用 τ 表示，是个量纲一的量.

反射率(reflectance or reflectivity) 一个电磁波传播的物理量. 即物体表面的反射波强度与总入射波强度之比. 即由反射引起的出射度与入射到物体表面的辐照度之比. 常用符号 ρ 表示，其数值在 0～1之间. 不同材料的表面具有不同的反射率，同一材料对于不同波长，其反射率也不同，这种现象称选择性反射. 反射率数值还与物体表面状况、其周围的介质及入射角有关. 对一特定表面，常称反射比. 在给定波长处单色光的反射率，称光谱反射率. 遥感影像分析的根据之一就是不同地物在不同谱段具有不同的反射率. 用同一地区的几张不同谱段的影像进行比较可提高判读的准确性. 半球反射率的定义与反射率相同.

发射率(emissivity) 亦称比辐射率. 一个电磁波辐射的物理量. 即是物体的辐射出射度与同温度的黑体的辐射出射度的比值，常用 ε 表示. 黑体的 ε 恒等于1，其他物体的 ε 皆小于1. 非透明物体的发射率与其吸收率相等. ε 是波长的函数，如各种岩浆岩，随岩石中的二氧化硅含量的降低，其发射率的谷值向长波方向移动. 遥感中可据此来区分岩性. 物体的 ε 值的大小还与物体表面状态有关. 例如，常温下白云岩磨光面的比辐射率 $\varepsilon = 0.929$，而粗糙面的 $\varepsilon = 0.958$.

大气窗口(atmospheric window) 若干特殊电磁波谱段. 即指电磁波在大气中传播时透过率比较高(或大气吸收作用相对比较弱)的波段. 常用的大气窗口有：

可见光及近红外谱段：0.30～2.5μm；

红外谱段：3.5～4.2μm，8～14μm；

微波波段：1.4mm，3.5mm，8mm 附近的波段，以及波长大于厘米波的微波波段.

遥感器只能运用大气窗口的电磁波段获取地面信息.

飞行姿态(flight attitude) 一种航空航天技术术语. 即对飞行器(包括航空和太空飞行器)在空中或太空中飞行姿态的描述. 通常用三个角度，即俯仰角、偏航角和滚动角来表示飞行姿态. 这些姿态角一般根据以飞行器重心为原点的直角坐标系的三条轴来定义. 通常将轨道平面内沿轨道的切线上的飞行方向定义为 x 轴的正方向；在过飞行器重心且平行于过星下点的地球切面的平面内、垂直于 x 轴的过飞行器重心的直线定义为 y 轴，按右手系定义其正方向；过飞行器重心的地球法线方向则为 z 轴. 由此，飞行器围绕 z 轴的偏转角称为航偏角，围绕 x 轴的偏转角称为滚动角，围绕 y 轴的偏转角称为俯仰角. 就太空飞行器而言，为保持姿态稳定，通常采用三种方式：

1. 自旋稳定. 太空飞行器绕轴自转，自转轴对太空某点取向固定，使飞行器保持姿态稳定.

2. 三轴稳定. 三轴多指以太空飞行器滚动、俯仰、偏航命名的直角坐标系的三个坐标轴. 飞行器本体不旋转，依靠姿态控制器使其取向稳定.

3. 重力梯度稳定. 在地球重力场作用下，转动物

体的转轴可以逐渐达到平衡状态，与重力梯度方向一致，即同当地铅垂线方向一致，从而保持太空飞行器姿态稳定.

地球同步轨道(earth synchronous orbit) 一种特殊的卫星轨道. 即卫星运动方向和角速度与地球自转方向和角速度分别相同的轨道. 这种轨道上的卫星的周期和地球自转周期一致. 如轨道平面与地球赤道面的倾角为零，即形成“地球静止轨道”. 如取倾角 $i°$，则轨道的地面轨迹在南北纬 $i°$以内成“8”字形. 地球静止轨道只有一条，高度约 35800km. 处于这样高度的卫星，能观察地球 40%以上范围. 因此，从理论上讲，只要在太空中均匀间隔布设三颗卫星，就能基本覆盖全地球. 这种轨道适于对地球长期地连续监视. 某些预警卫星和气象卫星就采用这种轨道，但对地球南北极那样的高纬度地区不能观测(参见“大地测量学”中的“同步轨道”).

数据传输分系统(data transmission subsystem) 卫星向地面发送数据的设备子系统. 即按数据格式将遥感数据(包括图象)或其他信息实时地或经星载磁带机记录后向地面数据接收站发送的分系统. 遥感系统中的下行通道传输是宽带数据传输. 它只传输遥感系统获取的原始数据或图象，并不兼顾外差. 目前一般采用 S 波段、X 波段和 K 波段，数据传输速率为每秒数十万比特至数十兆比特.

星历数据(ephemeris data) 亦称星历表. 一种星体轨道参数表. 即用列表数据说明每隔一定时间某星体预定所在位置，或每隔一定时间某人造卫星预定所在位置.

星历表(ephemeris) 即“星历数据”.

校准数据(calibration data) 一种遥感技术数据. 即与遥感器和辐射源的光谱特性、几何特性有关的各种测量值. 校准数据可利用具有固定能量分布的辐射源(如标准灯)、温度板或几何测试图表等来获取. 它是一种临时数据.

地面站(ground station) 卫星或航天系统的一个组成部分. 即设置在地球上的进行太空通信的地面设备. 遥感地面站的任务主要是监控卫星运转情况、接收遥感和遥测数据以及对信息进行数据处理和贮存等. 美国“陆地卫星”地面站由五部分组成：

1. 地面测控中心. 监视和跟踪卫星的运转情况，指挥星上遥感器工作和发送遥感数据.

2. 地面接收中心. 接收卫星发回的遥感图象数据和转发的遥测数据.

3. 地面数据处理中心. 将地面接收站的高密度数字磁带转换成计算机兼容的磁带和 70mm 胶片，并对上述磁带和胶片进行加工处理，晒印成粗制相片或精制相片.

4. 图象分析中心. 对数据处理中心产生的磁带、胶片进行影像增强、自动识别分类、产生专题图.

5. 综合数据库.

图象预处理(image preprocessing) 图象数据处理中的一个过程. 即提供应用或分析之前对图象的一种基础处理. 通常是指复原处理. 例如，卫星图象的预处理包括三方面内容：

1. 辐射亮度校正. 消除太阳角与大气引起散射的影响、遥感器的非线性和非一致性失真、胶片显影的非线性等，即进行各种校准和定标处理.

2. 多谱段图象的配准、调整.

3. 几何校正(参见“几何配准”和“几何校正”).

几何校正(geometric correction) 遥感图象处理中的一个重要内容. 即对遥感图象的几何畸变进行校正的工作. 几何畸变有两类：

1. 由于遥感器自身的结构性能非理想化或其指标偏离标称数值所引起，如摄影机标定主距与实际主距不等、物镜系统的光学畸变差、扫描型遥感器扫描运动的非直线性等.

2. 由于遥感器的位置、姿态和目标物所引起，如遥感器高度与姿态角变化、大气折光、地球旋转、地球曲率、地形起伏等.

几何校正的原理就是把有畸变图象各元素变换到所选定投影图象的相应位置上. 几何校正的方法有光学机械法和数字纠正法. 当前光学机械法主要应用于静态遥感器图象的几何校正，数字纠正法主要应用于动态遥感器图象的几何校正. 数字纠正法包含的内容有图象像元空间位置变换和像元灰度值的重抽样. 该法根据纠正数学模型又可分为参数法和非参数法. 参数法是对影像的变形规律或遥感器在飞行时的位置和姿态进行模拟和计算，求出变换参数，再用这些参数改算所有点. 此方法中多采用多项式或共线方程进行计算. 非参数法是基于控制点上的几何变形值推算出图象上其余位置的变形，可用最小二乘滤波法.

几何配准(geometric registration) 遥感图象处理中的一项内容. 即使不同图象能够完全叠合的处理过程. 即将不同时间、不同谱段、不同遥感器系统所获得的同一地区的图象(数据)，按同名点在位置上和方位上完全叠合的操作. 各个谱段的图象几何叠合称为光谱配准. 不同时间的图象几何叠合称为空间配准. 几何配准的目的是提高计算机自动分类的精度，监测地物特征的变化及其他图象处理需要. 几何配准的基本方法是求得各同名像点的相关值.

光谱配准(spectral registration) 见“几何配准”.

空间配准(spatial registration) 见“几何配准”.

大气校正(atmospheric correction) 清除卫星遥感图象在大气传输中所引起的退化因素的一种图象处理方法.大气的退化包含:对从地面向上辐射的电磁波信号的吸收、散射作用,大气湍流对辐射信号的扰动及大气本身的辐射成为地面辐射信号的噪声干扰.常用的大气传输模型为

$$L_i = G \cdot E \cdot T \cdot R_i + G \cdot L_p$$

或

$$F'(u,v) = T(u,v) \cdot F(u,v) + N(u,v),$$

式中 L_i 为遥感器输出信号,G 为遥感器增益,E 为太阳辐照度,T 为大气透射率,R_i 为地面反射率,L_p 为大气路径辐射率,$F(u,v)$,$F'(u,v)$分别为地面辐射信号及遥感器输出信号的傅里叶变换式,$T(u,v)$ 为大气传输特性,$N(u,v)$为噪声的傅里叶变换式.大气校正的方法一般有:

1. 通过对大气传输特性或参数的实际测量获得 E,T,L_p,$T(u,v)$,$N(u,v)$参数,从而确定 R_i 或 $F(u,v)$.

2. 通过地面与卫星对地物的同步测量,使用统计回归方程确定大气校正量 $L_i = a_i + b_i \cdot R_i$,式中 a_i,b_i 为回归系数,L_i 为卫星测量结果,R_i 为地面实测地物反射率.同步测量的 L_i,R_i 为已知数,由回归方程可确定 a_i,b_i,由此外推遥感图象的其他像元点值.

3. 在一定条件下应用同一像场的多谱段图象分布确定校正量.

图象复原(image restoration) 亦称图象恢复.图象处理的重要内容之一.指校正图象退化效应,将图象重建成接近于或完全无退化的理想图象.一般复原过程包括退化效应模型化、复原处理器设计、空间域或频率域上执行运算等步骤.复原问题并无通用解法,需根据不同的复原任务(如去除图象模糊、校正几何畸变、减弱噪声效应等)对退化模型和复原方法做特定的选择.例如,用卷积模板法去除线性运动模糊,用会话型复原法去除图象上周期噪声,用维纳滤波复原法去除卫星图象上的薄云覆盖等.

图象恢复(image restoration) 即"图象复原".

图象增强(image enhancement) 图象处理的内容之一.应用计算机或光学设备改善图象视觉效果的处理.处理模型是根据人眼对光亮度观察的特性确定的.目的是提高图象的可判读性.图象增强处理的内容包含反差增强和滤波.反差增强处理在于改善图象上类别的判读效果;滤波处理是为提取或抑制图象的边缘和细节特征、消除噪声等.常用的反差增强方法有对比度扩展、彩色增强、多谱段图象组合与变换.滤波增强分为空间域滤波和频率域滤波.

图象变换(image transformation) 图象处理的内容之一.泛指按一定规则从一帧图象加工产生另一帧图象的处理过程.在数字图象处理中,指用于复原、增强、分类、编码的各种变换.这时输入图象(处理对象)和输出图象(处理结果)是由离散数据组成的数字图象或其显示图象.图象变换可分为点变换、谱段变换或空间变换(局部变换).其数学运算包括通常的函数计算、统计分析基础上建立的变换函数、一维线性变换、二维线性变换或其他更复杂的运算.通过变换,使得图象便于识别或有助于进一步处理.

彩色合成(colour composite) 图象增强的处理方法之一.是将多谱段黑白图象变为彩色图象的处理技术.一般是三色合成,也可两色或四色合成.合成方法有:直接使用光学方法和使用计算机的数字处理方法两种.前者将一组黑白透明图片放入配有特定的红、绿、蓝三色滤光片的光学系统内,投影到同一屏幕上,使图象精确重合,形成彩色图象.数字处理合成法是令三图象的像元亮度值变换为红、绿、蓝三基色的彩色编码去控制彩色显示设备,形成彩色图象.其颜色组合是任意的.根据合成图象的彩色与实际景物自然彩色的关系,可分为真彩色合成与假彩色合成.当三图象的信号与自然景物三色光对应,合成结果为真彩色或逼近真彩色;如不对应或缺少某一色光时,则为假彩色.真彩色合成使图象目标颜色符合人眼观察自然景物的习惯;假彩色合成则由于原图象不受限制而形式多样,合成结果彩色鲜明、特征突出.

模式特征(pattern feature) 简称特征.模式识别中的一个术语.即图象模式识别中,一模式的各种测量值之一或这些测量值的某种数学变换.一般包括灰度、纹理、形状等信息.在遥感图象中,往往指遥感器通道,或地物影像的反射率测量值,或其数学变换值.如陆地卫星 MSS 图象 4 个谱段的数值构成 4 个特征,其谱段的线性组合构成线性组合特征.与某一模式有关的特征数目,定义为特征维数.

模式识别(pattern recognition) 亦称图形识别.一种图形和图象处理方法和技术.即对图形和图象进行区分、计数、定位、分类和解释的技术.其处理对象可以是空间的几何形态,也可以是代表某种特征的一组数据,均称模式.模式识别有一系列方法和技术,如试探法、统计法、语法结构法和直方图分析、空间频率分析、掩模相关等技术.识别过程分为预处理、特征萃取(选择、分类)和判定.实施的手段有光学处理和数字处理.

图形识别(pattern recognition) 即"模式识别".

监督分类(supervised classification) 一种先验(已知)类别标准的分类方法.对于待研究的对象或区域,先用已知类别或训练样本建立分类标准,而

后对样本的观测数据进行分类,是一种受控(监督)的信息类别识别过程.如遥感多谱段数据的判别分析方法.

非监督分类(unsupervised classification) 一种无先验(已知)类别标准的分类法.对于待研究的对象或区域,没有已知类别或训练样本作标准,而直接依据样本观测资料的内在联系进行分类.例如,遥感多谱段数据的集群分析方法.

判读(interpretation) 指从图象获取信息的基本过程.即根据各专业(部门)的要求,运用判读标志和实践经验,或借助各种技术手段和方法对遥感图象进行研究,判别出所需要的地物或测算出某种数量指标的过程.例如林业判读,是在航空相片上先识别森林类型,然后测算森林分布面积和木材蓄积量.根据判读信息的特征,可分为定性判读和定量判读;根据判读的技术和方法,可分为目视判读和自动化判读;根据判读内容,可分为一般判读和专题判读等.

目前,测绘、地理、林业和军事部门均使用此术语;地质等部门习惯用"解译"或"判释"等.

图象处理系统(image processing system) 用于进行图象处理的计算机系统.分为数字处理系统和模拟处理系统两大类.数字处理系统包括主机及各种外部设备(磁带机、磁盘机、图象显示器、图象扫描仪、打印机、绘图仪、键盘及终端监视器等).软件包含用于图象处理的应用软件,通常为人机对话式或菜单式.有些系统连结功能较强的图象处理机作为中央处理机的外部设备.

遥感图象制图(mapping from remote sensing images) 处理和判读遥感图象(如航空相片、红外影像、雷达影像、多谱段和多时相卫星影像等)、制作或更新地图和专题图的技术.用于判读和制图的遥感图象需经过处理(包括辐射校正、几何校正、图象增强、信息提取等),以达到符合判读和制图的要求.遥感制图的成果包括地形图、各种专题地图和影像地图等.

工程测量学

工程测量学(engineering surveying) 测绘学中的一个分支学科.它是测绘学在国民经济和国防建设中直接应用的一门技术学科.主要研究在工程、工业和城市建设以及资源开发的各个阶段进行地形和有关信息的采集和处理,工程建筑物的施工放样,设备安装测量,工程的变形监测、分析和预报的理论、技术和方法,以及研究对测量和与工程有关的信息进行管理和使用的学科.

古代在治理洪水、兴修水利以及土地划分中就进行了工程测量.司马迁在《史记》中有夏禹治水(公元前 21 世纪)时进行勘测的记载.北宋时沈括为了治理汴渠,测得"京师之地比泗州凡高十九丈四尺八寸六分".在国外,工程测量可追溯到古埃及的大金字塔建设(公元前 27 世纪)和幼发拉底河与尼罗河流域的土地边界测量(公元前 14 世纪).

工程测量按工程建设的规划设计、施工兴建和运营管理三个阶段分为工程勘测、施工测量和变形监测.其内容包括:提供模拟或数字的地形资料;进行测量及其有关信息的采集和处理;施工放样;大型精密设施的安装和调试测量;工业生产过程的质量检测和控制;各类工程建筑物、矿山和地质病害地带的变形监测、机理解释和预报;工程测量专用仪器的研制;与研究对象有关的信息系统的建立和应用等.

工程测量按所服务的工程种类,又分为建筑工程测量、水利工程测量、军事工程测量、海洋工程测量、精密工程测量、城市测量、矿山测量、三维工业测量、线路测量、桥梁和隧道工程测量以及工程摄影测量等.

工程测量学与测绘学的其他分支学科有着密切的关系.在建立工程测量控制网方面,涉及大地测量学和测量平差的内容.为了经济合理地提供各种比例尺的地形图,还要涉及摄影测量与遥感和地图制图学等.为了施工放样和变形监测的需要,要研制各种专用测量仪器,要涉及光学、机械学和电学、电子学的内容.此外,还要涉及有关的工程知识、测量数据处理、控制网优化设计、变形观测数据的几何分析、物理解释和变形预报、测量自动化以及建立工程信息系统等,还要涉及数学、物理学、计算机科学的有关知识.

工程测量控制网(engineering surveying control network) 进行工程测量的一种基础设施.即为工程建设测图、施工、管理的需要,在一定区域内布设的平面和高程控制网.它具有提供基准、控制全局和限制测量误差积累的作用,是各项测量工作的依据.二维平面控制网和一维高程控制网一般分别独立布设,有时也一次布设成三维控制网.平面控制网大多采用局部坐标系.布设专用控制网,往往有其特殊要求.对于控制区域较大的网,应与国家控制网联测.根据情况,既可采用分级布网,也可采用一次性全面布网.布设时应按最优化理论确定最佳方案.平面控制网一般采用三角、导线、三边和边角测量方法建立;高程控制网用几何水准测量和三角高程测量方法建立;用电子速测仪同时作三角高程测量或三边测量可建立三维控制网.目前利用 GPS 技术建立各种工程测量控制网日渐增多.

控制网质量准则(quality criterion of control network) 亦称控制网质量指标或质量标准.是控

制网优化设计的依据和目标.它包括精度、可靠性和经济性准则.对变形监测网还有响应变形的灵敏度和分离变形的可区分度.精度准则一般由控制点坐标估值的协因数矩阵出发,定义一些纯量精度标准.在优化设计时要按网的精度要求构造一个准则矩阵(参见“准则矩阵”).网的可靠性表征的是发现粗差和抵抗粗差影响的能力.灵敏度和可区分度是指变形监测网能发现某一变形模型的能力及其与另一变形模型或平差中观测值的粗差相区分的能力.控制网的经济性准则用于构造反映建网经费的目标函数.

控制网优化设计(optimal design of control network) 进行控制网设计的一种理论和方法.根据给定的要求和实际条件,采用最优化理论,确定测量控制网最优布设方案的方法.早在1868年,赫尔默特(Helmert,F.R.)就在“合理测量的概念”博士论文中,首次论述了最佳权分配问题.在中国20世纪50年代,曾研究过基线网优化设计.格拉法伦(Grafarend,E.W.)对大地测量控制网优化问题进行了广泛的研究,提出了四级优化分类概念,现已被国际测量学界所公认:零级设计是网的基准选取问题;一级设计是网的图形结构问题;二级设计是观测值的权的分配问题;三级设计是旧网的改进和加密问题.随着计算机科学和计算方法的发展,控制网优化设计在理论研究和实际应用方面都得到不断的发展.控制网优化设计可分为解析法和模拟法两种,它们都是以控制网的精度、可靠性、费用以及变形监测网的灵敏度为依据(参见“控制网质量准则”)的.解析法通常是在约束条件下,求极值或通过矩阵运算直接求解;模拟法则是通过计算机快速计算出控制网的精度、可靠性等指标,与设计值比较,测量工程师根据专业知识和经验,对布设方案进行修改,直到获得一个优化的、切实可行的方案.解析法和模拟法各有优缺点,一般地,解析法的结果是最优的,但不一定切实可行;模拟法的结果不一定最优,但却是可行的.由于模拟法可采用通用平差程序,并以人机交互的形式进行,所以它的适应性更强,更易于推广.

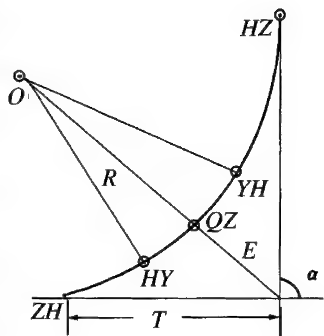
准则矩阵(criterion matrix) 控制网设计中给定的精度标准.在控制网优化设计时,按一定要求,人工构造的未知数向量(网点坐标或其函数)的协因数矩阵.构造的方法分相关函数法(或称协方差函数法)和直接配置法两种.相关函数法是将网点坐标向量的误差向量视为空间的一个随机过程,对它导出理想化(如均匀性和各向同性)的协方差函数,进而构造相应的准则矩阵.直接配置法是按给定的点位误差椭圆元素构造准则矩阵,或用奇异值分解法来缩减协因数矩阵的特征值,使所构造的准则矩阵满足特定的要求.用解析法进行控制网优化设计时,常根据准则矩阵或其逆阵的元素建立函数模型,采用最小二乘法或数字规划法逼近准则矩阵.

建筑工程测量(building construction survey) 工程测量的种类之一.是指建筑工程在设计、施工和运营期间的测量工作.设计阶段主要是为总平面图设计(包括初步设计和施工图设计)提供大比例尺地形图资料.施工阶段的测量工作有建立施工控制网、建筑物放样、竣工总平面图测绘或编绘(参见“竣工测量”).施工控制网除采用三角、边角和导线网外,常常布设成由正方形或矩形格网组成的建筑方格网,方格网的边与主要建筑物的轴线平行,其边长视建筑物大小而定.方格网点通常采用归化法测设.建筑物平面位置放样,一般按先总体后局部、先轴线后细部的原则进行.对于大型工业企业,还要在施工控制网下布设厂房控制网,用以放样厂房细部.高程放样根据水准控制点用水准测量方法进行,建筑物各部分标高通常都是相对于“零”标高线(建筑物底层室内地坪设计高程线),向上为正、向下为负.建筑物运营期间的测量主要是变形观测,它对监视建筑物的安全和验证设计理论具有重要意义(参见“变形观测”).

建筑方格网(square control network) 见“建筑工程测量”.

建筑限差(tolerance on building) 关于建筑物误差的一种度量.建筑物竣工时其实际尺寸相对于设计尺寸的容许偏差.它是由施工误差和测量误差所引起的.测量误差包括控制测量误差和放样误差.在给出建筑限差后,可以根据误差分配原则(如等影响原则、忽略不计原则、按比例分配原则等)确定控制测量和放样的精度,以此为依据进行控制网设计和放样设计.

线路测量(route survey) 工程测量的种类之一.它是指在铁路、公路、高压输电线、长距离渠道和管道等线型工程(统称线路)中勘测设计、施工建设和运行管理阶段所进行的测量工作.其中,铁路线路工程最具代表性,对测量的要求也最高.铁路勘测分初测和定测两个阶段.初测是按踏勘提出的不同方案,沿线路进行平面、高程控制测量和测绘带状地形图,供在带状图上设计线路的走向和坡度以及编制初步设计书;定测是将所设计的线路放样到实地,主要内容有中线测量、曲线测设和断面测量.中线测量分放样和中桩测设,前者是将图上定线的各交点间的线段测设到地面上,后者是在线路中线上测设百米桩,加桩和曲线主点桩.曲线测设的任务是在两相邻直线段间测设所设计的曲线,通常是一条带缓和曲线的圆曲线(见图).缓和曲线是连结直线和圆面线的过渡曲线,其曲率半径从无穷大变到圆曲线半径 R.曲线主点为:直缓点 ZH,缓圆点 HY,曲中点



QZ，圆缓点 YH 和缓直点 HZ. 当圆曲线半径 R、缓和曲线长度 l_0 和转角 α 已知时，可计算出曲线要素：切线长 T、外矢矩 E、曲线长和切曲差，并据此测设出曲线主点和细部点. 断面测量包括纵断面测量和横断面测量，前者是测量线路中线桩地面高程，后者是测量垂直于线路中线方向的地面起伏.

中线测量(center line survey)　见"线路测量".

曲线测设(setting out of curve)　见"线路测量".

断面测量(profiling survey)　工程测量中的一种常用测量. 沿某一方向或轴线测量地面(含水底)起伏的工作. 分为纵断面测量和横断面测量. 纵断面测量沿线路中线进行，在确定出中线桩的里程和高程后，可绘制纵断面图. 为了显示地形的变化，高程比例尺通常比水平距离比例尺大 10 或 20 倍. 纵断面图主要为设计坡度和计算土石方量之用. 横断面测量以中线桩为准，在与中线垂直的方向上分别测量两侧变坡点至中线桩的平距和高差，并据此绘制横断面图. 横断面图主要用于设计线型工程的横断面形状、计算土石方量、放样边坡和布设其他建构筑物. 纵、横断面图除在地面实测外，也可根据地形图或数字地面模型绘制. 对于河流纵断面图，一般是根据水下地形图、河道横断面图、水文水位资料进行编绘，断面系沿深泓点(河床最深点)绘制.

桥梁工程测量(bridgework survey)　工程测量的种类之一. 各种桥梁工程在勘测设计、施工建设和运营管理期间所进行的测量工作. 在勘测设计阶段，要为桥址选线提供所需的各种比例尺的地形图、水下地形图和水文测验资料；在施工建设阶段，为确保桥梁墩台的定位精度和桥梁的架设质量符合设计要求，需要进行桥址控制测量、各种施工放样和竣工测量；在运营管理阶段，为监视桥梁的安全运行，为桥梁维护和设计提供可靠的依据，需要定期地进行桥梁变形观测、数据处理和稳定性分析.

隧道工程测量(tunnel survey)　工程测量的种类之一. 隧道工程(含地铁和隧洞工程)在设计、施工和运营阶段所进行的测量，包括大比例尺地形图测绘、地面和地下控制测量、联系测量、隧道开挖过程中的施工测量和贯通测量、隧道断面竣工测量、根据需要对隧道及附近地面的变形观测等. 地面、地下控制测量以及联系测量要根据隧道的贯通精度进行设计. 除采用常规的地面测量方法外，GPS 在地面控制测量中逐渐得到应用；陀螺经纬仪在联系测量和地下导线测量中也被广泛采用；在用联合掘进机进行开挖时也采用激光导向测量系统.

竖井联系测量(shaft connection survey)　将地面和地下控制联系到同一坐标系中的方法. 在各种隧道工程建设和地下采矿巷道开挖过程中，通过竖井将地面和地下的平面和高程控制联系在同一坐标系统中的测量工作. 它包括平面联系测量(通常称竖井定向测量或竖井定向)和高程联系测量(或称竖井高程传递)两部分. 竖井定向有一井定向、两井定向和陀螺经纬仪定向(参见"陀螺经纬仪定向测量")等方法，用悬挂吊锤或用激光铅直仪投点，通过竖井向地下传递坐标和方向；竖井高程传递通常采用悬挂长钢尺或长钢丝配合水准仪的方法进行，也可用电磁波测距仪直接测量竖直距离.

陀螺经纬仪定向测量(gyrostatic orientation survey)　亦称陀螺定向. 测定真北方位角的一种方法. 用陀螺经纬仪测定待定边的真北方位角时，由于陀螺仪具有定轴性和进动性，在地球自转的影响下，陀螺旋转轴会在测站天文子午面两侧作往复的简谐摆动，通过对陀螺轴摆动的观测，可以在陀螺经纬仪上测出待定边的真北方位角，其定向精度，按一次定向中误差来分，可高达±3″. 用陀螺经纬仪定向可以控制方向的误差积累，常用于联系测量和地下导线测量.

逆转点法(reversal points method)　亦称跟踪法. 陀螺定向的一种传统方法. 定向时，使陀螺经纬仪的照准部连续地跟踪陀螺转子轴的摆动，记录其逆转点在水平度盘上的读数，取 5～7 个读数，采用舒勒平均值法，或托马斯平均值法，或相关平差法，可得真北方向在度盘上的读数，若再观测某一待定向方向的角值，即可得到该方向的方向值. 该法属于无扭观测，要使照准部处于跟踪状态，对观测人员在技术上要求较高.

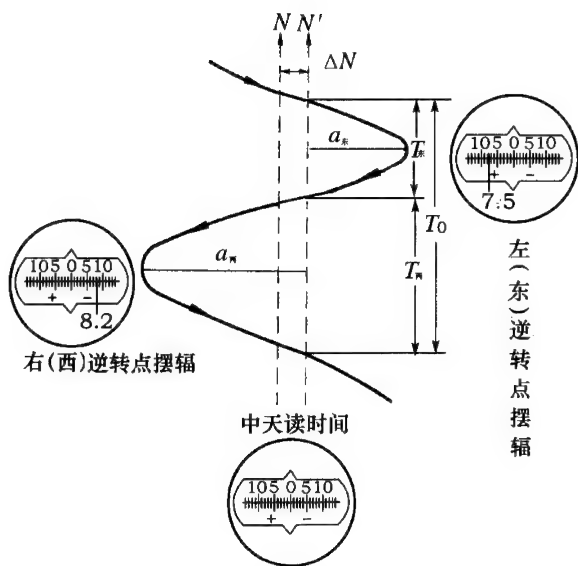
跟踪法(tracking method)　即"逆转点法".

中天法(transit method)　陀螺定向的方法之一. 它是通过测定陀螺灵敏部经过中天(分划板零刻画线)的时刻 t 和摆幅 a，确定真北方向. 作业时将陀螺经纬仪望远镜固定在近似北方向 N' 上，连续观测并记录陀螺转子轴经过中天的时刻和逆转点的摆幅(见图)，由中天时间计算左右摆动的时间差 Δt，由公式

$$\Delta N = c \cdot a \cdot \Delta t$$

计算近似北方向的改正数 ΔN(式中 c 为比例系数，可由观测数据或由摆动周期计算)，从而得到真北方向在度盘上的读数

$$N = N' + \Delta N.$$



若再观测某一待定向方向的角值，即可得到该方向的方位角．与中天法属于同类的还有记时摆幅法和改化摆幅法等方法．

水利工程测量(hydraulic engineering survey) 工程测量的种类之一．水利水电工程在勘测设计、施工安装和运行管理阶段所进行的测量工作．在勘测设计阶段需要测绘各种大、中比例尺地形图，进行线路测量、纵横断面测量以及回水边界线测量等，同时还要进行包括水位、水深、流速、流向、悬浮质和推移质测量的水测验；在施工安装阶段主要是建立施工控制网，进行施工放样、工程施工进度管理、设备的安装测量以及竣工测量；运行管理阶段的测量工作是对建筑物及病害地带的变形观测．对于水工建筑物来说，变形观测特别重要，包括的内容也很丰富(如大坝的外部和内部变形观测，船闸、高边坡监测和库内滑坡监测等)，观测的时间特别长，精度要求很高(参见“变形观测”)，有些工程从勘测设计开始就要进行某些变形观测．

港口工程测量(port engineering survey) 工程测量的种类之一．港口设施的新建、改建和维护修复工程(统称港口工程)在勘测设计、施工兴建和运营管理阶段的测量工作．勘测设计阶段要进行控制测量、港区陆地和水下大比例尺地形图测绘、底质探测、水文观测和港区资料调查等．施工阶段的测量主要有施工控制网的布设、建筑物平面位置和高程的放样、竣工测量以及施工中的变形观测等．在运营管理阶段，要对港口重要设施，如码头、防波堤进行水平位移、沉降、倾斜和裂缝等变形进行观测，用以监视变形的发展和掌握变形的规律，并提出防治的措施，确保建筑物的安全，还要为今后的设计积累资料．

水下地形测量(underwater topographic survey) 工程测量中的一种特定测量．测量江河、湖泊、水库、港湾和近海水底点的平面位置和高程，用以绘制水下地形图的测绘工作．主要内容是在陆地建立控制网和进行水下地形测绘．水下地形测绘包括测深点定位、水深测量、水位观测和绘图．测深点定位的方法有断面索法、经纬仪或平板仪前方交会法、六分仪后方交会法、全站式速测仪极坐标法、无线电定位法、水下声学定位和差分 GPS 定位法等．水深测量采用测深杆、测深锤和回声测深仪等器具．水底高程是根据水深测量和水位观测成果计算，最后用等深线(或称等高线)表示水底的地形情况．

军事工程测量(military engineering survey) 工程测量的种类之一．它是指在军事工程建设的勘测设计、施工兴建和运营管理阶段所进行的测量工作．目的是为各种军事工程建设提供测量数据和地形图，保障工程建设按设计竣工和进行有效的管理．按军事工程的种类可分为军事道路测量、坑道测量、军用机场测量、靶场工程测量、军用港口工程测量以及其他特殊工程测量．按测量精度分为普通和精密军事工程测量两种．

在古代，中国秦汉时期修建的万里长城、古罗马构筑的兵道和公元前 218 年欧洲修建通往意大利的“汉尼拔通道”，都进行了地形勘测和定线测量；在两次世界大战中修建的大量军事工程(如法国长达 390km 的马奇诺防线)、各种通信线路、管道、军用机场和军港等，都需要进行测量工作．随着科技的飞速发展，军事工程对测量要求的精度和可靠性都愈来愈高，如远程武器和卫星发射、地下核试验、大型抛物面天线和洞库工程等，对点位和安装精度的要求高达亚毫米量级，这就形成了精密军事工程测量．目前的趋势是向自动化方向发展．

军事工程测量在勘测设计、施工兴建和运营管理阶段的测量工作与一般工程测量的内容基本相同，包括测绘 1∶500～1∶5000 大比例尺地形图，在测图、施工和变形监测中所进行的控制测量，施工放样测量，竣工测量以及沉降、位移和挠度观测等变形测量．

靶场工程测量(range engineering survey) 军事工程测量中的一种测量．在武器试验区进行的测量工作．分为常规武器靶场测量和导弹试验发射场测量．前者主要包括场区控制测量、靶道线上中间点测量和桩位点测设等．导弹试验发射场测量主要包括场区的地形测量，对发射阵地、地下控制中心和测试厂房等建筑物的放样，竖井测量和跟踪观测设备的安装测量，施工过程中重要建筑物的变形观测．此外，还有一些特殊要求的测量，如导弹瞄准测量(确定发射基准边的方位和为瞄准提供射击方向，标定射向等)、短边方位角传递、校零和标校测量等．

机场测量(airport survey)　工程测量的种类之一.修建飞机机场中的测量工作,包括机场的勘测定点测量、净空带及场道测量、机库及地下油库测量和线路测量.勘测定点又分初测和定勘两个阶段,前者的任务是选出若干个场址,定勘是选定场址,其主要工作是测绘各种比例尺地形图和控制测量.净空带测量在于测定净空带内影响飞行安全的障碍物的位置,绘出净空带平面图及纵断面图.场道测量一般要布设方格网,测绘场道 1∶2000 的地形图和进行施工放样.机库及地下油库测量包括地面控制测量、施工放样和库区地形图测绘.线路测量有拖机道、进场公路、场外排水线路和输油管线路测量,主要内容是纵横断面测量和绘制纵横断面图.

地质测量(geological survey)　工程测量的种类之一.为布设物探测网和测绘编制地质图所进行的测量工作.按不同的目的,可分为地球物探测网测量、区域地质测量和专门性地质测量,后者如第四纪地质测量、水文地质测量及构造地质测量等.

工程摄影测量(engineering photogrammetry)　见"非地形摄影测量"和"近景摄影测量".

三维工业测量(three dimension industrial suvey)　工程测量的种类之一.利用光机电和摄影测量技术在航空航天工业、汽车和船舶制造等现代工业中为进行产品质量控制而提供在特定坐标系下的精确三维坐标的测量工作.三维工业测量采用经纬仪(含全自动化)系统、摄影测量系统和非常规量测系统(包括三维坐标量测仪、激光干涉仪、长度或角度传感器、全息术比长器、量测显微镜和断面投影器等).在近距离下,上述系统均可达到 10^{-5}的相对精度,测量过程逐步向实时处理和自动化方向发展,并开始广泛地用于各种现代工业之中.

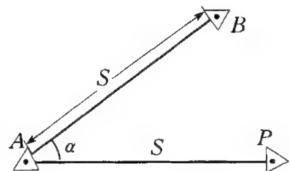
安装测量(installation survey)　工程测量中的一种测量.工程建设,特别是大型特种工程中对工艺设备及其构件的定位、安装、检核和调校所进行的测量工作,包括建立安装测量控制网、工艺设备定位和安装测量、检核测量和调校测量等.安装测量控制网具有构网坚强、图形划一、点的密度均匀等显著特点.主要有直线型、中心辐射型和直伸环型控制网等典型图形.通常,工艺设备的定位和安装要求的精度很高,不能直接用放样来保证,检核和调校测量常常要在不关闭设备和在高温、高压、带电和辐射等条件下进行,因此要求采用精度高、速度快、能遥测和自动控制的仪器和方法.除常规的精密测角、量距和水准测量仪器和方法外,需要采用特殊的方法和研制专门的仪器来保证控制网的精度和提高相邻部件的相对位置精度.采用的仪器有精密因瓦线尺和杆尺、尼龙丝准直和激光准直系统、激光干涉仪、测微准直望远镜、液体静力水准仪等.方法包括精密基准线法、引张线法、精密投点法、平滑测量法、测支距法和三角形测高法以及近景摄影测量法等.

施工测量(construction survey)　工程测量中的一种重要测量.它是指工程建筑物在施工兴建阶段所进行的测量工作,主要包括建立施工测量控制网、在局部区域测绘大比例尺地形图、建筑物放样、工艺设备及其构件的安装测量、施工过程或建成后的验收检核测量、竣工测量以及在施工期间进行的变形观测等.

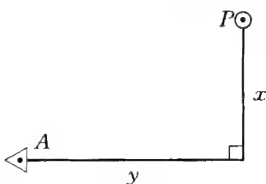
施工放样(construction setting out)　亦称放样或测设.施工测量中的一项工作.是按设计和施工的要求,将设计图上建筑物的空间位置和形状在实地标定出来的测量工作.点的标定称为定位,轴线标定称为定线.平面位置放样的方法有极坐标法、直角坐标法、正倒镜投点法、交会法及归化法等.其中交会法又包括前方交会、方向线交会、轴线交会和距离角度交会等.高程放样主要采用几何水准测量方法,有时也采用钢尺直接丈量竖直距离和三角高程测量的方法.

测设(setting out survey)　见"施工放样".

极坐标法(method of polar coordinate)　测定或放样点位的一种方法.它是根据一个方向和一段距离来确定点的位置.如图,在已知点 A 安置仪器,以另一已知点 B 定向,测得角度 α 和距离 S,则可得待定点 P 的坐标.若 P 点是放样点,则可由 A,B,P 的坐标计算出角值 α 和距离 S,同样可以在 A 安置仪器,通过角值和距离计算出放样点 P 的坐标.极坐标法是测量和放样中使用最广泛的一种方法.

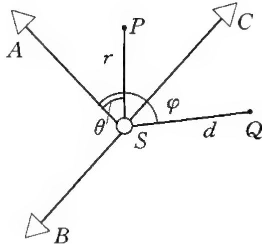


直角坐标法(method of rectangular coordinate)　测定或放样点位的一种方法.它是用纵、横坐标确定样点位置的方法.如图,首先在点 A 确定一条控制线,沿控制线量出横坐标 y,然后在垂直于控制线的方向量出纵坐标 x,则可得到 P 点位置.直角坐标法既是测量的一种方法,也是放样的一种方法,它只需量距离测设直角,在地籍测量和厂房施工放样中经常采用.用于铁路线路曲线测设时,又称切线支距法.

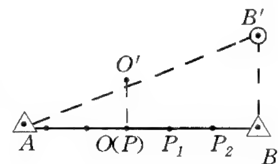


自由设站法(free station method)　确定测站位置的一种方法.将经纬仪或速测仪放置在某一点 S,采用后方交会原理,通过观测到两个以上已知点的方向、距离和高度角(若单纯的方向后方交会,则

至少需要三个已知点)，可以计算出自由设站点 S 的坐标. 然后就可以用任一已知点定向，按极坐标法测量任意点 P 的坐标，或者根据放样点 Q 的坐标计算放样元素进行放样(见图). 因此，自由设站法在测图和放样中都同样适用. 采用袖珍计算机编程，可以方便地计算设站点坐标、测量点坐标或放样点的放样元素.



正倒镜投点法(double reversing method) 放样直线或直线上的点的一种方法. 如图，用正镜(盘左)和倒镜(盘右)照准 A 点后，再倒转望远镜投点的方法. 即将经纬仪逐渐趋近地安置在 A,B 两已知点的连线上，然后用正倒镜照准 A 点，再用倒转望远镜投点的方法放样 AB 方向线上的点 P. 具体作法是：将经纬仪大致安置在 AB 中间的 O 点上，按正倒镜投点法得 B' 点，由

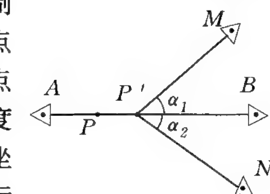


$$OO' = \frac{AO}{AB} \cdot BB'$$

计算出 O' 向 AB 连线方向移动的距离(AO,AB,BB' 可测得)，最后可将经纬仪安置在 AB 方向线上. 再用正倒镜投点法放样连线上的点 $P_1,P_2,\cdots$.

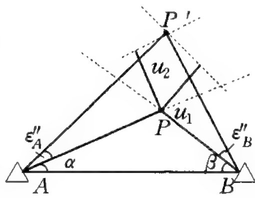
方向线交会法(directional line intersection method) 测定或放样点位的一种方法. 它是利用两条互相垂直的方向线相交确定点的位置. 方向线可根据施工控制网点设置，通常由建筑方格网和厂房控制网发展，用拉钢丝的方法放样细部点.

轴线交会法(axis-line intersection method) 放样已知直线上的点的一种方法. 如图，根据轴线端点 A,B 和轴线两侧的控制点 M,N，放样轴线上的点 P. 将仪器安置在靠近 P 点的轴线上 P' 点上，测定角度 α_1 和 α_2，可计算出 P' 的坐标，与设计放样点 P 的坐标比较，即可将 P' 移至设计位置 P. 此法常用于水利枢纽工程的施工放样.

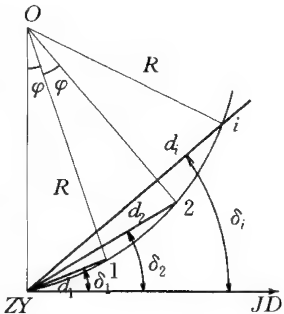


归化法(reduction method) 精确放样点位的一种方法. 根据已知点和放样点的坐标反算放样元素. 在已知点上架设仪器，采用某种方法先放样出放样点的近似位置得过渡点，然后精确测量过渡点的坐标. 根据过渡点的精确坐标和放样点的设计坐标，计算过渡点到放样点的改正数. 最后从过渡点出发，根据改正数将过渡点精确归化到设计位置. 用这种方法可以精确放样点位，也可以精确放样距离、角度和直线.

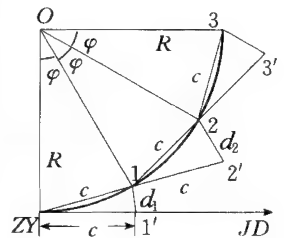
前方交会角差图解法(forward graphic intersection method with angular difference) 快速而精确地放样点位的一种方法. 如图，由控制点 A,B 和放样点 P 的坐标反算放样元素. 按前方交会原理进行放样点的初步定位，得 P' 点，然后在 A,B 上实测 α,β 角，根据实测角与理论角值之差 ε_i''，计算由角差所引起的放样点的横向偏差 u_i，将 P' 改正到设计位置 P. 该法常用于桥墩中心的放样.



偏角法(method of deflection angle) 曲线测设中放样点位的一种方法. 如图，根据切线和弦线之间的夹角 δ_i(称偏角)和弦长 d_i 交会的方法来测设曲线上的点位. 在铁路线路的曲线测设中常采用此法放样圆曲线和缓和曲线. 偏角 δ_i 可以根据圆曲线半径 R、曲线长度和缓和曲线长度为引数查表得到. 图例表示用偏角法放样圆曲线.



弦线偏距法(method of chord deflection distance) 曲线测设中放样点位的一种方法. 是把圆曲线上相邻两点间的弦长延长一倍，与计算得的偏距 d 一起，用距离交会法测设曲线上的点. 如图所示，由 c 和 φ 可计算出 d_1 和 d，在 ZY 和 $1'$ 可交会出点 1，延长至 $2'$，则在 1 和 $2'$ 可交会出点 2. 该法有时用于曲线隧道的开挖中.



竣工测量(finish construction survey) 工程测量中的一种测量. 是指工业企业建设和工程建筑物在竣工验收时的测绘工作. 对于主要建筑物的轮廓点、地下管线的转折点、道路交叉点、窨井中心等重要地物细部点，要用解析法测定坐标；对于主要建筑物室内地坪、上水道管顶、下水道管底和道路变坡点高程，采用水准仪测定；一般地物地貌则按地形测图要求测绘. 竣工测量的主要成果是竣工总平面图和特征点的坐标、高程明细表. 竣工总平面图除进行必要的野外实测之外，通常根据设计资料和施工测量资料进行编绘.

变形观测(deformation observation) 亦称变形监测或变形测量．工程测量中的一种重要测量．为保证工程安全，对建筑物及其地基和工程的地质病害地带，由于荷载和地质条件变化等外因作用而引起的动态空间变形，必须进行周期或持续测量．变形观测的目的：监视被测对象的安全情况，研究变形的规律及引起变形的原因，进行变形预报，以便采取措施防止变形可能造成的破坏或灾害．变形观测数据处理成果还可用来检验设计理论和计算方法是否正确．

变形观测包括水平位移观测、沉降观测、挠度观测、倾斜观测、摆动观测和裂缝观测．变形观测的方法有：精密水准测量，液体静力水准测量，地面摄影测量，精密准直测量，常规的方向、角度、距离测量，各种传感器测量以及 GPS 测量等．不同的变形观测项目通常还要求采用专用的仪器和特殊的方法．

变形观测数据处理是变形观测的重要内容，它包括变形监测网优化设计、观测数据的质量评定、观测值之间的相关性分析、观测值粗差和系统误差剔除、变形的几何分析、物理解释和预报等．

变形观测要求的精度很高，通常要求观测引起的误差应小于允许变形值的 1/20～1/10，有时甚至要求采用当前测量技术所能达到的最高精度．此外，要求进行周期性的重复观测，综合应用各种观测手段，采用严密的数据处理．在制定观测方案、进行监测网优化设计和进行数据处理时，需要多学科的配合．可以说，变形观测是处于测绘学、地球物理、岩土力学、土木工程等学科之间的一门边缘学科，也是工程测量学的重要组成部分．

变形监测(deformation monitoring) 即“变形观测”．

变形测量(deformation measurement) 即“变形观测”．

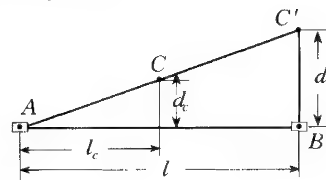
基准线法(reference line method) 亦称准直法．变形观测中测定水平位移的常用方法，也被应用于安装测量．设 A,B 为基准线的两个端点，通过它们的铅直平面称基准面，为确定 A,B 两点间一系列点相对于基准面的偏离值而进行的测量工作称准直测量，其方法称基准线法．按所使用的仪器划分，又分为视准线法、激光准直法和引张线法．视准线法是将经纬仪安置在 A 点，标牌安置在 B 点，由经纬仪的视准面形成基准面．视准线法又分为测小角法和活动觇牌法．前者是用经纬仪精确测定基准面方向与测站点到观测点的视线方向之间的小角，从而计算出观测点相对于基准线的偏离值；活动觇牌法则是利用活动觇牌上的标尺，直接读取偏离值．激光准直法又分激光经纬仪准直和波带板激光准直(参见“波带板激光准直”)，激光经纬仪准直是用激光束代替经纬仪视线．引张线法是通过拉直钢丝的竖直面作为基准面，主要用于大坝的水平位移观测(参见“引张线法”)．当基准线很长时，为了获得较高的精度，可采用分段基准线法．

准直法(alignment method) 即“基准线法”．

引张线法(method of tension wire alignment) 变形观测中测定水平位移的一种方法．利用一根拉紧的不锈钢丝建立基准面来测定观测点的偏离值的方法．引张线装置一般设在水工大坝的观测廊道内，由端点、观测点、测线(不锈钢丝)和测线保护管组成．为解决引张线垂曲过大的问题，在其间常设置若干浮托装置．测点上的偏离值可采用读数显微镜读取标尺上的分划或通过位移传感器自动测定．假设端点是固定不动的，而各观测点上的标尺或传感器是与坝体相连的，则偏离值的变化就代表该点所代表坝段的位移值．引张线法不受大气旁折光影响，一次观测的中误差可高达 0.03mm．

波带板激光准直(laser alignment with zone plates) 安装测量和变形观测中的一种典型方法．用激光干涉原理建立基准线，在基准线两端点 A,B 上分别安置激光器点光源和光电探测器，在观测点 C 上安置波带板，根据激光的特点和光波干涉原理，通过波带板的激光束在光电探测器上成像为亮点或十字线，可测出 AC 连线在 B 点处相对于基准面的偏离值 d，从而计算出 C 点对于基准面的偏离值

$$d_c = \frac{l_c}{l} \cdot d.$$



(见图)波带板激光准直常用于变形观测和设备的安装测量，其精度可达 10^{-6}，若安置在真空管道中，可达 10^{-7}的精度．

变形的几何分析(geometrical analysis of deformation) 变形观测数据处理的一种方法．提供变形体在外因作用下变形的几何形态和时间特性．分为静态模型、运动模型和动态模型．前两种模型一般分别用于变形监测网的参考网和相对网，采用多元假设检验方法对观测结果进行分析，以确定点的位移以及变形的速度和加速度等时间特性．动态模型往往是分析变形体在动荷载或周期性变化的荷载作用下的动态变形，主要根据变形观测的时间系列找出变形的频率和振幅．变形几何分析过程一般包括模型初步鉴别、模型参数估计和模拟统计检验及其最佳模型选取三个步骤．

变形的物理解释(physical interpretation of deformation) 变形观测数据处理的一种方法．确定变形体的变形与引起变形的原因之间的关系．通常

采用统计分析法(亦称回归分析法)和确定函数法.前者是通过分析变形和外因之间的相关性,建立荷载与变形间的数学模型;后者则是利用变形体的物理性质、材料的力学性质以及应力应变间的关系来建立变形模型,模型为位移场的微分方程,在实际工作中一般采用有限元法计算.统计分析法要利用已有的观测资料,当具有大量观测数据时,效果较好,故具有"后验"性质.确定函数法在变形观测初期也能建模并进行变形预报,有"先验"的性质.两种方法各有优缺点,且互为补充,所以综合运用两种方法较为理想,即采用回归分析和确定函数综合分析法.

精密工程测量(precise engineering survey) 工程测量中的一个重要分支.大型特种工程在设计、施工、设备安装调校以及运营管理过程中的高精度测量工作.它的特点是:极高的精度和可靠性要求,快速、遥测和自动化等.例如,在环形粒子加速器工程中,对于长达数十公里的环形轨道,要求两相邻电磁铁安装的径向相对误差不超过

$$\pm(0.1\sim0.2)\text{mm},$$

安装调校等作业中的测量工作往往要在高温、高压、带电和辐射条件下进行.精密工程测量的数据处理需要用到现代平差理论和方法,如粗差探测和剔除、方差和协方差分量估计、模型误差检验、动态变形分析以及按精度、可靠性和灵敏度准则进行网的最优化设计等.

精密工程测量技术主要有精密方向和角度测量、精密距离和水准测量、精密准直测量、精密投点和定位测量等,其精度越来越高.例如,真空波带板激光准直测量的精度可达 10^{-7},高精度电子经纬仪几乎达到了角度、方向测量的极限精度,双色电磁波测距仪的测距精度可达

$$\pm(0.1\ \text{mm}+0.1\text{ppm}\cdot D),$$

而双频激光干涉仪可达到 0.5 μm/m 的计量精度,采用 CCD 技术将把精度提高到纳米级.各种专用测量仪器和自动化观测系统在精密工程测量中广泛应用,电子测量技术、自动控制技术和计算机技术的飞速发展大大地改变了精密工程测量的面貌.

城市测量(urban survey) 与工程测量密切相关的测量分支.为城市建设的规划设计、施工兴建和运营管理所进行的测量工作.主要内容包括城市控制测量、城市地形图和地籍图测绘、各种专题图编绘、兴建市政工程和城市改造中的施工放样以及重要建筑物和地带的变形观测等.

城市控制测量需满足测绘最大比例尺(通常为1∶500)地形图或地籍图和市政工程施工放样的精度要求.除采用常规的地面三角测量、边角测量和导线测量方法建立控制网外,到 20 世纪 90 年代,主要采用 GPS 测量技术重建或改造城市测量控制网.城市地形图的比例尺按规划设计的需要一般为

1∶500~1∶25000,

分实测和编绘两种.为了保持地形图的现势性,需要经常修测或重测.采用的方法有航空摄影测量和地面模拟或数字化测量等.城市地籍图的比例尺与建筑密度有关,市区为

1∶500~1∶1000,

郊区为

1∶2000~1∶10000,

地籍图的现势性要求更强.地形图和地籍图的公共部分是基本平面图,其成图逐渐向数字化方向发展.各种城市专题图都以基本平面图为底图进行编绘.由城市基本图图形资料和各种专题图有关资料生成的城市基础地理信息系统,对城市的规划、建设和管理具有十分重要的意义.市政工程施工放样包括城市道路、桥梁、地铁、地下管网和各种公用建筑物在建设施工时的放样,以及拨地测量与规划红线放样等.变形观测主要对高大和重要建筑物进行,有的城市还需进行城市地表沉降观测.

矿山测量(mine survey) 工程测量的种类之一.在矿山建设和开采过程中,为获得各种矿图图件和解决各种几何问题所进行的测绘工作.主要内容有:建立矿区地面控制网,测绘比例尺为

1∶500~1∶5000

的矿区地形图;对于露天矿,要测绘采掘平面图;进行矿山施工测量,包括地面土建工程测量,井下巷道掘进时的测量,将地面与地下控制网联系在同一平面坐标系统和高程系统的测量(参见"竖井联系测量"),为确保掘进的巷道按设计要求准确衔接的测量;为研究由于地下开采而产生的岩层和地表的移动规律而进行的测量(参见"地表移动观测")以及矿体几何图绘制(参见"矿体几何图")等.

地表移动观测(ground displacement observation) 矿山测量中的一种测量.为研究由于地下开采而产生的岩层和地表移动规律所进行的测量工作.主要内容有:在采区地表沿矿体走向和倾向埋设由测点组成的观测线,定期进行平面位置和高程测量,通过对不同地质采矿条件的大量观测和数据处理,确定地表移动的变形大小和分布规律,为合理开采提供可靠数据;通过向岩层内部打深钻孔设点,观测岩层的内部移动;监测地表建筑物、铁路路基的移动和变形以及露天矿边坡的稳定性等.

矿体几何图(geometrical map of mineral deposit) 矿山测量中的一种图.描述矿体、岩层和结构面几何形状和物理化学特征的图件.为此,要测定表征形状和物理化学性质的各种标志值,对所得到的数据进行数理统计分析,用点和注记精确地描绘在图上,根据实际情况进行内插和外推,给出等值线来

反映标志值在平面和空间上的变化规律. 表示矿体形态的,称为构造图;表示矿石品质的,称为性质几何图.

地籍测量(cadastral survey) 亦称不动产测量. 与工程测量密切相关的测量分支. 确定和调查土地及其上附着物的权属、位置、数量、质量和利用现状,为土地的管理和利用所进行的测绘工作. 地籍一词源于拉丁文 catastrum,意为人头税登记. 地籍测量的发展包括建立征税地籍、产权地籍和多用途地籍三个阶段. 在中国,从公元前 2000 年殷、周时期的"井田制",到明、清的"履亩丈量"、"土地清丈"都说明了地籍测量的悠久历史和重要作用. 地籍测量的对象是地块或宗地及其上的建筑物,统称为不动产. 地籍测量的主要内容包括测绘不动产地籍图和建立不动产地籍册,作为土地管理部门或房地产部门办理土地登记、发放土地所有权或使用权证书的依据. 同时地籍测量也包括土地分等定级、地产价格评定、土地利用调查和面积量算等内容. 地籍测量资料除用于征税外,在法律、规划、管理和经济等方面都具有重要意义,它具有法律效力. 目前,地籍测量逐渐向建立多用途地籍的方向发展.

地籍测量的方法有地面测量和航空摄影测量,包括地籍图根控制网和碎部测量两大部分. 由于地籍图的现势性要求很强,地籍图上的地块和房屋等地籍要素的基本平面图形数据具有相对的稳定性,因此,采用数字化成图和数据库管理,实现地籍图册的自动化建立、修测和管理是地籍测量的必由之路. 地籍数据库也是城市地理信息系统或土地信息系统的基础.

不动产测量(real estates survey) 即"地籍测量".

地籍图根控制测量(cadastre mapping control survey) 地籍测量中的一种基础工作. 即为测绘地籍图的控制网的加密测量. 为了满足测绘地籍图和进行经常性的地籍修测的需要,在国家基本控制网下用地面测量、GPS 测量或航空摄影测量方法进行. 可以采用分级布网或全面布网的方式. 控制点的精度和密度需满足规范的要求,每个点要埋设永久性标志,并按一定规则编号. 地籍图根控制点资料是不动产地籍的重要证明文件.

地籍碎部测量(cadastre detail survey) 地籍测量中的常用测量. 确定地籍细部点坐标的测量. 与地形碎部测量相似,但不测绘等高线. 测量的主要对象是行政区边界、地块或宗地边界、房屋建筑、固定地类界、道路和水系;其次有在地形和历史文化上有意义的对象,如墙篱、堤坝、桥涵、高压电杆、里程碑、纪念碑塔、遗迹等;此外还可根据用户要求测量路灯、路标、窨井、通风井、自来水阀门、花坛等对象. 测量的方法有直角坐标法、极坐标法和航空摄影测量法等. 目前,地籍碎部测量逐渐从模拟测图向数字化成图方向发展.

面积量算(measure of area) 地籍测量中的一个术语. 它是指对地块或宗地、房屋、土地利用种类的面积进行的测量和计算工作. 在土地征税、统计和地产价格评定等方面具有重要意义. 量算的方法有解析法和图解法. 前者是根据边界点的解析坐标按公式计算;后者包括用求积仪、格线板等工具在图上量算,还可用数字化仪与计算机相结合进行电算.

海洋测绘学

海洋测绘(ocean surveying and mapping) 测绘学的一个分支学科. 它是以海洋水体和海底为对象所进行的测量工作和海图编制工作. 表征海洋水体和海底性质的参量主要有水深、海底地貌、底质、海洋重力、地磁、海面形态、潮汐、海流、波浪以及海水的物理和化学特性等. 海洋测绘的基本任务是:准确地测量这些参量,并以其成果提供和应用于海上交通运输、海洋地球物理勘探和资源开发、海洋工程、国防建设、海洋环境保护,以及地球形状、海底地壳构造和运动的研究等领域. 海洋测绘的基本内容包括海道测量、海洋大地测量、海底地形测量、海洋重力测量、海洋磁力测量、海底底质探测、海面形态测量、海洋工程测量,以及其他专题测量,还要编制航海图、海底地形图、海面形态图、海洋重力图、海洋磁力图、海洋水文图以及其他专题图. 20 世纪 50 年代以来,海洋测绘在应用新技术和扩大研究内容方面取得了巨大进步,如多波束测深系统、机载激光测深系统和海洋遥感测深的试验和应用,卫星测高技术用于海洋大地水准面、重力异常、海洋环流、海洋潮汐等的探测和研究以及自动化成图技术等. 海洋测量工作当前已从测量航海要素为主,发展到测量各种专题要素的信息和建立海底地形模型的全部信息,进而建立各种海洋信息系统,特别是大型综合测量船的配备、获取全部信息的自动化采集数据系统和自动化成图系统,已使海洋测绘提高到了一个新的水平.

海道测量(hydrographic survey) 亦称水道测量. 海洋测绘的类型之一. 以保证航行安全为目的,包括对海洋、江河、湖泊等水域的水体和水下地形进行的测量和调查工作. 海道测量的基本任务是测量有关水域的制图要素、进行海区资料调查,为编制海图、编写航路指南和海洋科学研究提供资料. 海道测量的基本内容包括地形岸线测量、控制测量、水深测量、扫海测量、底质探测、水文观测、助航标志和海区资料调查等. 在港湾地区需测制大于 1 : 25000 比

例尺图;在距海岸 10 海里内的沿岸海区,则测制 1:25000至 1:50000 比例尺图;在距岸 10～200 海里的近海海域,则测制 1:100000 至 1:500000 比例尺图;在超过 200 海里以外的远海海域,一般只需测制小于 1:500 000 的比例尺图.

中国海道测量有悠久历史.早在《吴子》中已提出了水域深浅与军事的关系.宋朝已有测天定位、绳索测深和探测底质的方法.明朝航海家郑和七次下"西洋"期间,沿途进行测量水深、探测底质等工作,制作了世界最早的航海图集——《郑和航海图》.1922—1949 年间,中国海道测量局进行了一些海道测量工作,出版海图近 40 幅.1949 年以来,海道测量得到了很大发展,完成了中国海域基本测量工作,出版了大量海图和航海图书资料,制定了海道测量规范.随着科学技术的发展,海上定位手段已由光学、电子仪器发展到激光仪器和人造卫星定位技术.测深方法已由回声测深仪发展到多波束测深系统和遥感测深技术等.整个海道测量工作正向数字化、自动化方向发展.

水道测量(hydrographic survey)　即"海道测量".

回声测深(echo sounding)　测量水深的一种方法.它是利用声波传播往返时间差测定水的深度.通常选择频率在 1500Hz 附近的超声波,垂直向水底发射声信号,并记录从声波发射到信号由水底返回的时间间隔,通过模拟或直接计算,确定水的深度.按照这个原理设计的水深测量仪器称为回声测深仪.只能在一条测线上连续测深的多为宽波束的回声测深仪.而能在一定宽度的面上同时获得数十个水深点的称为多波束回声测深系统.由于后者具有覆盖面积大、效率高的优点,已在海上工程施工、重要航道的大面积精确测量以及绘制海底三维图等方面得到广泛应用.

多波束回声测深系统(multi-beam echo sounding system)　见"回声测深".

海底地形测量(bathymetric surveying)　海洋测绘的类型之一.是以测量海底起伏为主的测绘工作.海底地形测量通常由安装在船上的回声测深仪和侧扫声纳同时测定水深和水下地物大小和位置,也可采用多波束回声测深系统、海底摄影测量、机载激光测深和海洋遥感测深等方法.根据距海岸的远近,各测点的定位可采用光学定位、无线电定位、水声定位、卫星定位和组合定位等方法.在利用潜水船进行海底地形测量时,可采用惯性定位或利用海底控制点定位.其中以多波束回声测深系统与组合卫星定位系统相配合,由计算机实时处理绘出海底三维图形的方法发展最快.海底地形图在坐标系统、基准点、分幅、比例尺等方面应统一规定,并与同地区的陆地地形图相一致,以利于海、陆图的衔接使用.

海洋大地测量(marine geodetic survey)　海洋测绘的类型之一.它是指在海洋上,包括海面和海底进行的大地测量工作.它的基本任务是:为测制海图和海洋勘探以及工程测量提供基本的平面和高程(深度)控制;为海面和水下运载体确定精确的三维位置;为研究地球形状、大小、海底地壳构造和运动以及海域地震、潮汐变化等提供重要资料.海洋大地测量的基本内容包括在海域布设海洋大地控制网(主要是海底控制网)、海面和水下载体定位、海洋重力测量、海面地形测定、利用卫星测高资料和重力测量资料研究和确定海洋大地水准面、对海底地壳构造和运动以及海底地震进行监测等.

海洋重力测量(marine gravimetry)　海洋测绘的类型之一.是指在海洋范围内进行的重力加速度测定工作.在浅水海域通常使用海底重力仪,利用遥测装置在海面上进行观测.这种方法精度虽高但效率低.目前广泛使用安装在船上的各种海洋重力仪,在航行中进行重力测量.但航行中的重力仪会受到各种干扰,如厄特沃什效应的影响、水平加速度和垂直加速度的影响、交叉耦合效应(C-C 效应)的影响等,测量中必须消除或改正这些影响.海洋重力测量的精度,除受重力仪误差以及各种干扰影响外,很大程度上取决于舰船导航定位的精度,因此,进行海洋重力测量时,应采用高精度的导航定位系统.海洋重力测量资料是研究地球形状和地球内部结构、勘探海洋矿产资源、保障航天和远程武器发射等应用的重要依据.

GPS 海洋测量(GPS for marine survey)　海洋测绘的类型之一.它是指应用 GPS(全球定位系统)在海上进行的海洋测量工作.由于 GPS 能在海、陆、空上任何地点为载体全天候、全自动、快速连续、高精度提供位置、速度和时间信息,因此 20 世纪 80 年代以来,GPS 技术在海洋测量中得到了广泛应用.主要有:应用 GPS 载波相位相对定位建立岛礁海洋大地测量控制网,为领土和专属经济区划界、近岸海道测量、海洋工程测量提供控制基础;应用 DGPS(差分 GPS)技术建立以水下声基阵为标志的海底控制网,为水下潜航、监测海底地壳运动、深海资料勘查、水下施工以及海面载体动力定位提供控制基础;应用 GPS 伪距、载波相位、差分技术进行动态水深测量,绘制各种比例尺的水深图、海底地形图等基本图件;应用 GPS 测高数据确定海洋大地水准面等.目前,各种应用 GPS 技术于海洋测量的设备或系统已越来越多地投入使用,从而极大地提高了海洋测量技术的水平和效率.

扫海测量(sweeping survey)　海道测量中的一项重要工作.它是利用扫海设备对测定海区进行

探测，以查明该海区内或在海区所规定的深度上是否有航行的障碍物，包括确定航行障碍物的大小、性质和准确位置以及确定航道、港口或海区的最大安全通航深度．扫海设备可用硬式或软式机械扫海具，也可用海底地貌探测仪（亦称侧扫声纳）、多波束回声测深系统和海洋磁力仪等．扫海测量方法有定深扫海测量和拖底扫海测量两种．进行面扫测时扫海趟有效宽度的选取，应以最少的扫海趟覆盖扫测区为原则，相邻扫海趟之间要有重叠，以免发生漏测．

海底底质探测（seabed geology survey）　海洋测绘的类型之一．是指探测海底表层物质组成的工作．海底表层一般由陆源物质沉积物、火山沉积物、生物沉积物或暴露的基岩组成．在浅水海域可用采泥器或水铊进行底质探测，在较深海域，普遍使用浅底层剖面仪．底质探测一般与水深测量同时进行．底质点的密度，要根据需要和海底表层底质状况来确定．在一般海图上，以缩写注记的方法表示底质，也可编制成底质专题海图．底质资料是舰船锚泊、潜艇坐底、水中武器发射以及海底地壳构造和运动等的重要参考资料．

海洋遥感测深（ocean remote sensing sounding）　测定海洋深度的一种方法．利用航空或航天（如卫星）运载工具上的遥感测深系统测定海底各点在水面下深度的工作．按工作原理有机载激光测深系统、多光谱扫描测深系统和两者相结合的混合光谱扫描测深系统．机载激光测深系统是通过测定两束波长不同的激光束分别射向海面和透过海水到海底并返回所需的时间，来计算水深．该系统在测深精度和测量效率方面都比一般回声测深仪要高．但探测深度受激光器功率及海况条件限制，目前仅达数十米．多光谱扫描测深系统是根据不同光谱段渗透海水的能力不同，采用几个狭窄的不同波段的光谱带进行水下扫描，从而获得不同深度的断面图象．该系统有效测量深度在一般海况下仅达数米．混合光谱扫描系统既能通过激光脉冲测定水深值，同时又能通过扫描获得不同深度的断面图象，从而获得海底地形资料．因此，该系统适用于宽海域、分辨率要求高的机载水深测量．

机载激光测深系统（airborne laser sounding system）　见“海洋遥感测深”．

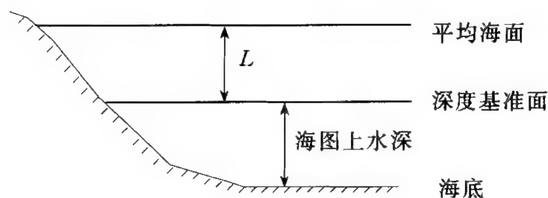
多光谱扫描测深系统（multi-spectral scanning sounding system）　见“海洋遥感测深”．

混合光谱扫描测深系统（mixed spectral scanning sounding system）　见“海洋遥感测深”．

水深测量（sounding）　水下地形测量的基本方法．它是测定水底各点平面位置及其在水面以下的深度，是海道测量和海底地形测量的基本手段．测深器具通常使用测深杆、水铊、回声测深仪、多波束回声测深系统和海底地貌探测仪等．所测得瞬时水面下的深度，经测深仪改正和水位改正，可以归算到由深度基准面起算的深度．

海面地形（sea surface topography）　描述海面稳态动力起伏的一个概念．它是指平均海面相对于某一等位面（大地水准面）的倾斜．平均海面可视为无干扰的稳态海水面，它相对于大地水准面的高，称为稳态海面地形，简称海面地形．对近岸海域海面地形的测定，可采用大地水准联测法，即以某一验潮站的平均海面作为高程起算的重力等位面（大地水准面），再以精密水准联测其他各站的平均海面高程，其差值即为海面地形．在大洋海域，可采用海洋水准方法或卫星测高方法测定海面地形．海面地形资料不仅用于研究海洋大地水准面，还能用于深海潮汐、海洋形态以及海流等的研究．

深度基准面（depth datum）　关于度量海水深度的一个依据．深度基准面是海图及各种水深资料中的所有深度的起算面．深度基准面通常取在当地



平均海面下深度为 L 的位置（见图）．确定深度基准面的原则是：既要考虑到舰船航行安全，又要考虑到航道的利用率．由于各国求 L 值的方法不同，因此所采用的深度基准面也不相同．中国在海洋、港湾和河口地区，主要采用理论深度基准面（即理论最低低潮面），它由 13 个主要分潮（M_2，S_2，N_2，K_2，O_1，K_1，P_1，Q_1，M_4，MS_4，M_b，S_a，S_{sa}）的调和常数按一定计算公式求得；在内河和湖泊采用最低水位、平均低水位或设计水位等作为深度基准面．

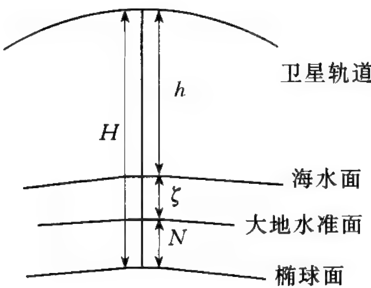
理论最低低潮面（lowest normal low water）　见“深度基准面”．

验潮站（tidal station）　记录潮位升降变化的观测站．设置验潮站的目的是了解海区潮汐性质及其变化规律．根据对验潮的要求和观测时间的长短，验潮站分长期验潮站、短期验潮站、临时验潮站和定点验潮站四种．长期验潮站又称基本验潮站，它长期观测所积累的资料是确定当地平均海面、理论深度基准面以及研究潮汐变化规律的依据．验潮站应选在能反映海区潮汐变化规律、远离河口、不受风浪影响且地质坚实的地点．潮位观测一般采用水尺或各种自记验潮仪，在远海域则可采用回声测深仪等．验潮站的水位观测值均从验潮站零点起算．

海洋工程测量（marine engineering survey）

海洋测绘的类型之一.在海洋工程建设勘测设计、施工和管理阶段所进行的各种测量工作.海洋工程主要包括海港、码头、堤岸、助航标志、养殖场、海底电缆、管道、海上钻井平台以及国防工程等.海洋工程测量的内容包括:勘测设计阶段的控制测量和陆岸与海底地形测量;施工阶段的施工测量和设备安装或复位测量;管理阶段的变形(或动态)观测和维修维护测量等.海洋工程测量的技术和方法与一般工程测量大体相同,对于水域中工程测量还广泛采用水声技术和方法.

海洋大地水准面(marine geoid) 根据地球重力场理论描述地球(海洋部分)形状的一个基本概念.它是大地水准面的海洋部分.它通常以处于流体静平衡状态的海水面(即平均海面)来表述.从理论上讲,平均海面并不是重力等位面,只是海洋大地水准面的一种近似.20 世纪 70 年代发展起来的卫星测高技术,已成为目前测定海洋大地水准面的最有效手段.见图,h 为雷达测高仪测得的卫星至海水面



间的高度.当卫星轨道参数已知并选定了参考椭球,则卫星至椭球面的距离 H 就可求得.在 h 中加潮汐改正,并顾及海面地形 ζ(其幅度为±(1～2)m),便可求出大地水准面差距

$$N(=H-h-\zeta),$$

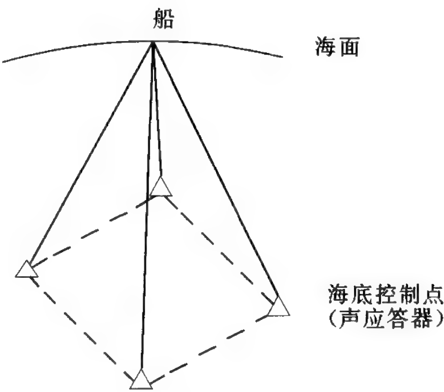
从而给出海洋大地水准面.

卫星测高(satellite altimetry) 卫星应用技术之一.它是利用卫星上的雷达测高仪测定卫星至瞬时海水面间的高度.雷达测高仪向海面发射脉冲波并接收返回信号,通过信号的时延可计算出卫星至瞬时海水面之间的垂直距离,称为海面高度,可由雷达波在海面射迹圆内所有反射波加以平均后求得.它消除了海面短波波动的影响.1992 年,美国和法国联合发射的 TOPEX/POSEIDON 卫星上的测高仪,测高精度已达±5cm.由于卫星测高覆盖面大、资料密集、数据精度高,已广泛应用于深海潮汐、海流、海面形态等的研究.同时,它也是研究确定海面地形和海洋大地水准面的有效手段.

平均海面(mean sea level) 描述海水面形态和变化的一个概念.它是指某验潮站一定时期内水位值的平均值.它代表该验潮站有效控制范围内海区的平均海水面.按连续验潮时间的长短,平均海面可分为日平均海面、月平均海面、年平均海面和多年平均海面.由于多年平均海面较好地消除了天文潮汐等周期波动的影响,可视其为处于流体静平衡状态的海水面.故在大地测量和海洋测量中都以多年平均海面作为高程起算面,又称基准面.例如中国规定以青岛验潮站的多年平均海面作为高程基准面.欧洲地区和美国分别以阿姆斯特丹和波特兰验潮站的多年平均海面作为高程基准面.多年平均海面受天文、海洋地球物理、气象等因素的影响,仍存在变化.研究海平面变化,可为研究全球气候变化、海洋与大气相互作用、海洋地球物理、环境保护以及沿海地区的经济发展,提供有价值的资料.

水位改正(correction of water level) 与测定水深有关的一个概念.它是把在瞬时海面上所测得的深度归算至规定深度基准面起算的深度所应加的改正.通常可根据一个、两个或三个验潮站的潮位资料计算.当测区在验潮站控制范围内,可根据一个验潮站潮位观测资料绘出水位曲线,从而求任意时刻的水位改正数;当测区在验潮站控制范围之外,在满足水位改正精度的条件下,可根据两个或三个验潮站的潮位观测资料用图解内插法、模拟法或解析法求水位改正,称水位分带改正.江河、湖泊水域,除河口区外,其水位主要随季节变化,水位改正所依据的水位站的布设、水位观测方法和资料处理等与海上均有所不同,故称江河水位改正.

海底控制网(submarine control network) 进行海洋测量的基础设施之一.是指布设于海底的大地控制网.海底控制网中的控制点是一些固设于海底的声应答器(见图).它们的三维位置以统一的大



地基准为参考,通常借助于海面上的测量船,利用岸上(如无线电岸台)或空间(卫星)的已知点采用声学测距技术予以测定.受海水中声波传播特性(有效作用距离)的限制,海底控制网多采取正三角形或正四边形布设,目前控制点间距离可达 15km 左右.海底控制网是海洋测绘中某些海洋工程建设和载体定位

的工作基础.

水文观测(hydrometry)　采集水体有关数据的一项工作.它是以江、河、湖、海的各种水文要素为主进行的观测.江河湖泊水文要素有水深、水位、流向、流速、流量、水温、冰情、比重、含沙量、降水量、蒸发量、水色、透明度、水的化学组成等;海洋水文要素有潮汐、潮流、波浪、海流、海水温度、盐度、海上气温、气压、风向、风速、浮游生物等.通常依一定条件在江河湖海的一定地点或断面上布设水文观测站,进行长期不间断的水文观测.各种水文观测资料经整理分析后,不仅是各种水文预报的依据,而且也是研究海床、河床、河岸变迁、海流、径流规律和进行各种水利工程、海岸工程设计计算以及编写航路指南等的重要资料.

无线电双曲线定位系统(radio hyperbolic positioning system)　海洋测绘中确定船位的一种方法.它是以距离差方式测定位置的无线电定位系统.由至少三个岸台和船台组成.双曲线定位的几何原理是:测定船台到两个岸台的距离差,可得到以这两个岸台为焦点的一条双曲线,两条这样的双曲线就可交出船位.按其工作方式可分为脉冲双曲线定位系统、相位双曲线定位系统和脉冲/相位双曲线定位系统.按作用距离可分为近程、中程、远程和超远程四种.以罗兰-C 为代表的远程系统,在 1000km 以内的定位精度约为 100m;近程系统作用距离一般为 250km 左右,定位精度可达 10m 左右.无线电双曲线定位系统具有全天候、连续、实时定位的优点,在海洋测绘中广泛使用.

海控点(hydrographic control point)　用于海洋测绘中的控制点.为海洋工程施工和进行大比例尺测图,在国家一、二、三、四等三角点和导线点间布设的加密点.这种加密点按精度分别称海控一级点、二级点和测图点.海控点坐标的测定可采用 GPS 测量、三角测量和导线测量方法.

海水中声速(submarine sound velocity)　描述海水中声传播特性的一个术语.它指声波在海水中的传播速度.声速大小随海水的温度、盐度和静压力的分布不均匀而变化,变化范围一般在 1420～1520m/s 之间.根据不同要求,现场海水声速可采用声速仪、声速表和具有不同精度的各种声速经验公式的计算来确定.

定位格网(positioning grid)　海洋测量中定位所用的一种辅助方法.指绘制在海图上用于舰船导航定位的几何格网.例如,辐射线格网是采用前方交会法定位,由岸上两控制点(经纬仪站)绘出的两簇辐射线构成;等距离格网是采用无线电测距系统定位,由两岸台为圆心的两簇等距离圆弧所构成;等角格网是采用六分仪后方交会法定位,由岸上控制点间连线为弦的两簇等角圆弧所构成;双曲线格网是采用无线电测距差系统定位,由一定台链配置的岸台为焦点的两簇双曲线所构成.绘制格网时,要求相邻双曲线之间的距离为 2～3cm,其他格网相邻直线或曲线之间的距离均为 1.5cm,而格网绘制线的误差均为±0.6mm.

辐射线格网(radial grid)　见“定位格网”.

等距离格网(equidistant grid)　见“定位格网”.

等角格网(equiangular grid)　见“定位格网”.

双曲线格网(hyperbolic grid)　见“定位格网”.

位置函数等值线(isopleth of position function)　海洋测绘中确定动点位置的一种方法.是指平面上位置函数为常数时动点的轨迹.在海洋测绘中,通常要求至少以两条位置函数等值线相交以确定定位点的平面位置.常用的位置函数等值线有距离函数等值线(它是一个圆)、方位函数等值线(它是一条直线)、角度函数等值线(它是圆弧)、距离差函数等值线(它是一条双曲线).通常只研究定位点附近的等值线就够了,并用通过定位点等值线的切线来代表其等值线.此切线称为定位点的位置线.

距离函数等值线(isopleth of range function)　见“位置函数等值线”.

方位函数等值线(isopleth of azimuth function)　见“位置函数等值线”.

角度函数等值线(isopleth of angle function)　见“位置函数等值线”.

距离差函数等值线(isopleth of distance difference function)　见“位置函数等值线”.

海图(chart)　海洋测绘工作的一个主要成果.是以海洋为主要描绘对象的地图.着重表示海洋区域的自然要素和社会经济要素.海图按其内容和用途一般分为两大类.数量最多的一类是直接用于航海定位、保证航行安全的航海图,包括海区总图、航行图、港湾图等.1982 年,第十二届国际海道测量大会通过了国际海图规范,国际海道测量组织各会员国将按照统一的规范编制国际通用的航海图,称为国际海图.另一类是不直接用于航海而以表示某些专题要素为主的专题海图,如海底地貌图、海洋底质图、海洋水文图、海洋重力异常图等.它们既是海洋专题调查的科学成果,也是进一步研究海洋、开发海洋资源必不可少的工具.

国际海图(international chart)　见“海图”.

专题海图(special topic chart)　见“海图”.

海洋磁力测量(marine magnetic survey)　海洋测绘的类型之一.是在海上进行地磁要素(磁偏角、磁倾角和磁场强度)测定的工作.早期海洋磁力测量主要采用饱和式磁力仪在飞机上或无磁性船上

进行.目前在近海已广泛使用能连续高精度测量的拖曳式船用质子旋进磁力仪.在远海使用质子磁力梯度仪.为监视磁暴和磁扰,消除日变,岸上应设日变站,对每天所测资料进行改正.海上磁力观测资料经日变改正、船磁改正以及正常场校正后,就可得到反映地壳岩石磁性差异的地磁异常,它是研究海底地质构造和地壳变化的重要资料.地磁要素资料在海道测量中还用来探明沉船、礁石、铁壳船的消磁、磁罗经的方位校正等.

地图制图学

地图制图学(cartography) 亦称地图学.测绘学的一个分支学科.是以地图信息传输为中心,研究地图的理论实质、制作技术和使用方法的综合性学科.按技术特点,地图制图学一般包括地图投影、地图编制、地图整饰、地图复制(制印)、地图应用等.现代地图制图技术能利用空间遥感技术获取地球和其他星体的信息来编制各种地图和天体图,并能以图解、数字或触觉形式组织、表达、传输和应用地理信息或空间信息,它与测量学、地理学、数学、印刷技术、电子技术、空间遥感技术、计算机技术、美学等有着密切联系.

地图学(cartography) 即"地图制图学".

数学地图学(mathematical cartography) 地图制图学中的一个分支和边缘学科.它研究数学在地图制图中应用的理论与方法.主要有:

1.地图投影的理论与方法.

2.数字地图制图过程中各种数学模型的建立,如地图图形向数字形式转换的数学模型、地图符号的数学描述与变换规则、地图要素空间分布与分析的数学模型(如曲面拟合、多变量统计分析等)、制图综合的数学模型(地图信息的抽象、选取和概括).

3.地图量算的数学方法.

专题地图学(thematic cartography) 地图制图学的一个分支.它研究专题地图的理论、编制技术与应用.其中主要有专题地图的内容实质、表示对象的分类分级和指标选取、专题内容的表示方法、主题要素与一般要素间的统一性与协调性、专题制图数学模型(区域群组和综合体的空间分布、以数量尺度反映的抽象分类、时间和空间信息的组合、要素间的相互分析、现象发展的动态分析与预测等)、专题地图符号系统与彩色设计、遥感资料在编图中的应用、专题地图分析与应用等.

地图编制(map compilation) 地图制图学的一个分支.是研究由制图资料制作地图的理论与方法,它的研究范围是从接受制图任务、收集和编辑地图资料到完成编绘原图的整个过程.具体包括地图编辑设计、编图技术方法、制图综合理论与技术等三个主要部分.通常,制图资料的比例尺、投影、内容同成图要求并不一致,需进行比例变换、地图投影变换和内容选取与概括.地图编制作业的一般过程为制图资料准备(选择、分析评价和加工)、研究制图区域与制图要求、拟订编图设计书、建立数学基础、转换地图内容、制图综合、地图整饰、图幅接边和审校及填写图历等.

地图整饰(map decoration) 地图制图学的一个分支.它是研究地图内容的表现形式和方法,主要是研究地图色彩整饰、线划整饰、符号设计、地貌立体显示、地图集的图幅编排与装帧等.其中,由编绘原图或地形原图按出版要求进行的线划整饰称为地图清绘.

地图复制(map reproduction) 亦称地图制印.地图制图学的一个分支.它是研究复制地图的各种工艺技术方法和理论,包括地图复照、地图制版和地图印刷三个主要部分.地图复制的作业过程(以平版制版印刷为例)通常包括原图复照、翻版、分涂、晒版、打样、样张审校和修改印版、印刷、成品检查及包装等工序.若原图为刻绘的,可省略复照工序.现代复制工艺已广泛采用电子分色制版、静电复印、缩微、无压力印刷等新技术.数字地图的复制则指计算机系统下不同存储介质间的地图数据转储,或经数-模转换分版输出胶片,直接用于晒版,然后打样、印刷,从而省去了复照、分涂等工序.

地图制印(map reproduction) 即"地图复制".

地图应用(map use) 亦称地图利用.是地图制图学的一个部分.它研究地图产品的特点和应用技术,按地图表示的各种现象的数量和质量特征及评价指标去研究其结构、相互关系、分布和动态变化.研究的方法有:

1.解析法.用统计或常规计数的方法按地图表达的信息做出数量和质量分析.

2.图解法.依地图的二维或三维模型制作剖面、断面块状图等.

3.图解解析法.进行地图量算,计算物体和现象的形态和结构的各种指标.

4.数学-地图模型法.按地图数据建立和分析其模型,并依据数学模型制作新的派生地图.

在科学、文化、教育、经济建设、国防等实践活动中,都要应用地图来进行作业和评价成果的可靠性和有效性.地图也用于历史学、社会学、人类学等学科中.

地图分析(map analysis) 地图使用中的一个主要方法与过程.是把地图作为客观现象的模型,对它进行科学分析与研究的工作.研究范围包括对地图表象的系统分析(研究它们的结构与相互关系)、

数量与质量特征的描述、现象的演变与发展. 分析方法主要有解析法、图解法、图解解析法、图形变换法等. 现代的地图分析方法建立在数学分析、统计学、信息论等学科理论以及图形测量、处理、变换等数字化技术基础上.

地图叠置分析(map overlay analysis)　对呈面状分布的各地理现象合成分析的一种方法. 是将同地区多种地理要素叠置分析,或将同地区不同时期的某种地理要素叠置进行多时相分析的工作. 在机助地图制图中,是将参与分析的若干个专题层面(分别代表不同地理现象或不同时期的某种地理现象)的多边形叠加(求交)形成新的多边形叠置图. 根据某种分析模型或专家系统,对各新多边形进行综合分析评价或确定其类别,最后产生分析报告或制作专题地图. 多边形求交和形成新多边形,可用矢量方法或栅格方法,亦可两者混合应用.

地图量算(cartometry)　地图制图学中的一个术语. 指在地图上测量计算各种要素的数量特征和精度分析. 其主要内容有:

1. 测量点位的地理坐标、平面直角坐标和高程.
2. 量算线状要素的长度、方向、曲率、挠率等.
3. 量算地表面积、坡度、坡向等地形参数.
4. 量算指定范围的体积或容积.
5. 根据地图投影计算加密经纬网,确定大圆航线和等角航线的位置等.

地图量算实质上是地图投影逆变换(由图上位置反算空间位置)与误差改正(成图作业误差、制图综合移位、图纸变形等),并计算各数量特征的过程. 常用的量算工具有曲线仪(里程计)、求积仪以及直尺、分规和各种模片等. 采用跟踪数字化仪和扫描仪进行测量,以及利用数字地图进行数据处理,能快速精确地获得各种量算结果. 这是地图量算技术的新发展.

地图传输(cartographic communication)　在地图制图过程中进行环境信息传输的理论. 是地图作者根据某一目的获取的原始环境信息进行符号化或运用地图语言的法则编码制成地图的过程. 地图是信息的载体和通道,用图者形成对环境的认识,即环境信息的恢复. 地图传输是 20 世纪 60 年代后期发展起来的一门地图制图学理论. 1972 年,国际地图制图协会设立了地图传输委员会,即第五委员会. 1976 年该委员会提出地图传输的研究方向是:地图传输的实质及理论基础、地图信息传递控制论模型、地图制图语言的实质及其语法和作为传输手段的地图的特性、空间信息纠正及信息转换的损失和增强、地图信息的感受及思想和心理生理过程等.

机助地图制图(computer-aided cartography or computer assisted cartography)　简称 CAC. 亦称计算机辅助地图制图或自动化制图. 地图制图的一种方法. 是用计算机及其外围设备作为制图的主要工具. 它采用数据库技术和图形数字处理方法,来研究地图信息的获取、变换、传输、识别、存储、处理、显示和绘图的应用学科. 机助地图制图的基本过程为数据获取(地图或影像数字化、数字测图、数据转储等)、数据处理(如几何改正、数据压缩、投影变换、拓扑结构化、制图综合、图形编辑、曲线光滑、曲面拟合、数字地图接边等)以及图形输出(二维或三维图形). 在地图或影像数字化(即模-数转换)时,需将图形或图象数字化成计算机能识别和处理的数据,主要有矢量形式和栅格形式两种.

计算机辅助地图制图(computer-aided cartography)　即"机助地图制图".

自动化制图(automated mapping)　即"机助地图制图".

地图(map)　测绘工作的一种主要成果. 用测绘技术把地球(或其他星球)上各种物体和现象表示在平面上的图形. 它是根据一定数学法则,使用地图语言,通过制图综合,以图解(数字的、触觉的)形式来表示各种自然现象、社会经济现象等地理信息或空间信息的图. 它反映各种自然和社会经济现象的空间分布、组合、联系及其随时间的变化与发展. 地图不同于其他图件的主要特征是采用一定的数学法则(地图投影方法),使地图具有可测量性,使用地图语言(地图符号及注记等)能直观地表示具体事物或抽象概念及有形或无形的地理现象,实施制图综合使地图具有一览性和反映事物的本质特征与内在联系. 地图内容由数学要素、地理要素和整饰要素三部分组成.

普通地图(general map)　地图的种类之一. 综合反映自然现象、社会经济现象一般特征的地图. 内容包括各种自然地理要素(水系、地貌、土质、植被等)和社会经济要素(居民点、交通线、境界线等),而不侧重表示其中某一种要素. 它可为经济、文化和国防建设提供资料,也可作为专题地图的地理底图. 不同比例尺的普通地图内容的详细程度、精度、制作方法各不相同.

专题地图(thematic map)　地图的种类之一. 根据专业需要着重选择一种或几种自然或社会现象作为主题要素予以表示的地图,如自然地图、经济地图、历史地图等. 其中主题要素表示要突出而详细,其他要素则视反映主题的需要作为地理基础选绘. 主题要素的表示法主要有质底法、范围法、点数法、等值线法、运动线法、统计图法等. 按地理分析的不同需要,专题地图可划分为现象分布地图(如人口分布、人口密度、地貌形态、地貌切割密度等)和分析评价的结论地图(如现象结构分析地图、现象过程关系

地图、动态预测地图、评价结论地图等)两大类.

数字地图(digital map) 地图的种类之一.是以数字形式表示的地图.即经过编码并以数字形式(矢量的或栅格的)将地图要素按一定数据结构存储在一定介质(如磁盘、磁带、光盘)上的数据集.它用属性、几何信息、关系来描述地图要素.其几何信息一般只表示地图要素的中心位置或中轴线,不直接存储符号图形,但经计算机处理后能将地图要素"符号化",并转换为模拟地图(以显示屏或纸张、胶片等为介质).数字地图具有不变形性、易编辑、易传输、容量(可存储的地图信息)大等优点.

模拟地图(analog map) 见"数字地图".

网格地图(grid map) 地图的种类之一.是一种以网格为单元表示的专题地图.即是将制图区域按一定大小的网格划分,以网格为单元存储专题要素,经计算机处理后仍以网格为单元表示专题现象的地图.网格尺寸取决于区域特点和制图目的.网格愈小,表示的内容愈详细,精度愈高,但存储量要增大.地形坡度、土地类型、土地利用等具有面状分布的专题要素,可以制成网格地图表示.网格地图除具有存储量小、便于计算机处理、输出较快的优点外,还便于区域综合分析和评价.

制图综合(cartographic generalization) 编制地图时处理地图内容的原则与方法.其实质就是在有限的地图信息载体上表示出制图区域的基本特征和制图现象的主要特点.通常表现为根据地图用途、比例尺和制图区域特点,对制图现象的选取、形状化简以及数量、质量特征的概括.从不同的策略和效果出发,制图综合分为:

1. 比例综合.选取和概括的程度主要取决于地图比例尺.

2. 目的综合.选取和概括不完全取决于比例尺,还要根据制图对象的重要性来确定.

3. 感受综合.从读者的实际感受效果出发进行选取和概括.

实际应用中,上述三种手段是配合运用的.制图综合不是对图形的简单缩绘或取舍,而是一个创造性过程.它表现为:对制图对象科学抽象及主次分明地反映内在联系和规律性,对不能以比例尺表示的重要物体采取夸张表示,正确处理地图内容的详细性与易读性及地图几何精确性与地理适应性的矛盾(通常采取变换表示方法、运用色彩效果、适当缩小或移位等措施).

统计图法(statistical diagram method) 制作专题地图的一种表示方法.它是在地理底图上或在地图图廓外配置图表以表示各区域单元内的统计资料的方法.它适用于多种专题要素分布现象的表示.分为:

1. 定位图表法.以统计图表形式在区域内定点表示特征现象和变化规律,如某一气象站上全年12个月的气温和降水量的变化曲线图表、风向风力图表等.

2. 分区统计图法.亦采用统计图表形式表示.但通常描绘在区域的适当位置,反映该区域内诸现象的总和、结构关系及动态变化.

3. 分级统计图法.按行政区划或经济区划用不同色调的色彩或形式和疏密不等的晕线表示现象相对指标的差异(级别),如人口密度、耕地与总面积的百分比等.

定位图表法(located statistical diagram method) 见"统计图法".

分区统计图法(areal division statistical diagram method) 见"统计图法".

分级统计图法(classified statistical diagram method) 见"统计图法".

等值线法(isoline method) 亦称等量线法.在制图中表示现象数量特征的一种方法.等值线是由制图现象中数值相等的各点连结成的连续曲线,如等高线、等深线、等温线、等压线、等磁偏线等.多用于表示连续分布且渐变的现象,如地势、气候等.根据需要,等值线的间距(值差)可设计成固定的或可变的.也可在相邻等值线间染不同的色彩以增加其明显性(如分层设色地图).

等量线法(isoline method) 即"等值线法".

分层设色法(hypsometric layer or hypsometric method) 制图中一种常用的方法.即用不同色相、色调表示地貌各高程层(带)的方法.地貌按高程划分为若干带,对每一带规定的具体色相和色调称为色层.用色的色阶表示各高程带,称为色层表.色层表有单色和多色两种.设色原则有:

1. 愈高愈暗.以光线直照原理愈陡愈暗为基础,色调随高程的增加而变深.

2. 愈高愈亮.以视觉感受为基础,认为读图时高地比低地离观察者更近,故高地用明亮色,低地用暗色.

此外,还有愈高愈饱和以及按自然设色等各种设色方案.实际应用时,往往综合采用上述设色原则以取长补短.

色层表(color layer table) 见"分层设色法".

晕渲法(hill shading) 亦称阴影法.地图上表示地貌的一种方法.用深浅不同的色调表示地面的起伏形态.德国学者于1716年绘制世界图时首先采用.按光源的位置分为直照晕渲和斜照晕渲;按色调分为墨晕渲和彩色晕渲.晕渲色调的浓度与地表坡面的受光量一般成反比,亦即光线方向与坡面法线方向的夹角其余弦越小,晕渲色调则愈浓而暗.

阴影法(hill shading) 即"晕渲法".

地图数据库(cartographic data base) 地图数字化后存储的一种形式.即是以地图信息为存储对象的数据库.是存储在一起的相关的地图数据集合,通常以图幅为单位建库.它是机助地图制图系统的核心,也是地理信息系统的重要组成部分.按地图数据形式分矢量数据库和栅格数据库.数据库中的地图数据主要由几何数据和属性数据(又称非图形数据)组成,采用一定数据结构来表示地理要素间相互联系的关系信息(主要是拓扑关系等空间关系).

地图数据库管理系统(cartographic data base management system) 亦称地图数据库软件系统.数字化制图软件系统之一.以地图数据为对象的专业数据库管理系统.因地图数据量巨大和空间相关的复杂性,使得通用数据库管理系统难以胜任对地图数据的管理,故需要专门研制地图数据库管理系统,或以某个通用数据库管理系统为基础开发成适用的地图数据库管理系统.

地图数据库软件系统(cartographic data base software system) 即"地图数据库管理系统".

制图专家系统(cartographic expert system) 计算机制图软件系统之一.应用人工智能技术进行机助地图制图的一组计算机程序和数据.它在机助地图制图过程中所解决的问题(主要是非数学描述的制图功能)具有制图专家水平.主要由地图制图知识库和推理机组成.知识库中存放经过研究和归纳的地图制图专家的知识和经验,其形式可以是谓词、规则,也可以是框架、语义,它们均需保持知识的完整性和协调一致性.推理机是一组程序,用来决定如何使用知识库和进行正向、反向或正反向推理.

地图数字化(map digitizing) 机助地图制图过程中,地图数据输入的主要方法.是将图形的连续模拟量转换成离散的数字形式,以便计算机处理.数字化方式分跟踪数字化和扫描数字化两种.前者获得矢量数据,后者获得栅格数据.

地理信息系统(geographic information system) 简称GIS.测绘信息系统之一.属于空间信息系统.为某种目标而建立,在计算机软件、硬件支持下,对有关空间数据按地理坐标或空间位置进行预处理、输入、存储、查询检索、运算、分析、显示、更新和提供应用、研究并处理各种空间实体及空间关系为主的技术系统."地理"在这里指"空间",表述信息的空间位置和关系.国际上该系统始于20世纪60年代初期,20世纪70年代渐臻成熟,20世纪80年代更有所突破.主要服务于资源与环境领域,在区域管理、规划和科学决策中也得到广泛应用.主要功能可概括为:

1.对空间信息进行多条件、多方式的查询检索,从而取得各种资源与环境要素的数据清单.

2.对空间信息进行统计分析的能力.

3.将查询检索和分析结果予以空间表达的能力,在不同的输出设备上,用图形和图象的形式,为用户提供直观的形象.

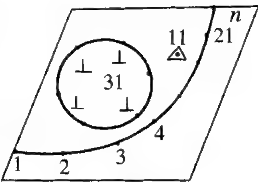
4.综合模拟分析和预测预报能力.

5.以专家经验和知识为基础的人工智能分析,形成专家系统,回答用户提出的咨询,提供决策方案.

矢量数据(vector data) 在计算机中表示空间数据的两种基本形式之一.由于空间数据具有属性特征(表示实际物体或目标、现象的性质)和空间特征(表示物体或目标的空间位置),所以在计算机中可用一对或一组坐标加上属性码(或称特征码)来表示空间数据,称该数据为矢量数据.

按照矢量数据结构,地图上基本图形要素可以概括为点、线、面三种元素及它们的组合,其表示方式示于下面的图和表中.利用全站型仪器野外测量、解析和数字摄影测量及地图数字化,可以直接获得这种矢量数据.

点	属性码	0011
	坐 标	(x_1,y_1)
线	属性码	0021
	坐标串	$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$
面	属性码	0031
	闭合坐标串	$(x_1,y_1),\cdots,(x_n,y_n),(x_1,y_1)$



矢量数据是一种面向现象、面向目标的数据表示方式,结构紧凑、冗余度低、几何精度好.但矢量数据的数据结构复杂,不便于进行空间分析和多边形叠置分析.目前,大多数地理信息系统中均采用矢量数据来记录、表示和存储空间数据.

栅格数据(raster data) 在计算机中表示空间数据的两种基本形式之一.将地图或影像规则地分割成极小的栅格单元,称为像元或像素(pixel).像元的大小取决于精度要求和原始图形或影像的分辨率.各像元按矩阵形式组织存储和构成关系,其位置由相应的行、列数来确定.对于地图而言,点状地物以其中心处的像元表示,线状地物则由中心轴线的像元连结构成,而面状地物则为其所覆盖的像元集合.通常用像元灰度值表示地理要素属性.

用栅格数据表示图形,数据结构简单,易于进行空间分析、多边形叠置分析和地理现象的模拟,便于

与遥感数据、数字地面模型数据和影像数据复合应用和分析，输入、输出速度快，自动化程度高.但栅格数据的数据量大、数学精度低、重采样较困难，而且面向现象和目标的识别效果也不如矢量数据.

栅格-矢量数据转换(raster-vector data conversion) 数据转换的一种方法.即矢量与栅格两种数据形式之间的转换技术.空间数据表示的两种方法各有优缺点和适用场合，因此需要根据使用目的进行栅格-矢量数据的转换.由矢量数据向栅格数据的转换一般比较方便.对于点、线目标，由其所在的栅格行、列数表示，对于面状目标，则需判定落入该面积内的像元.通常栅格(像元)尺寸均大于原来坐标表示的分辨率，所以若将栅格化数据再反转回去，则不可能达到原来矢量数据的精度.将矢量数据转化为栅格数据，主要用于空间分析、多边形叠置等.

栅格数据矢量化较为复杂，如果由一幅扫描的数字化地图来建立矢量数据库，则需要经过数字图象处理，如边缘增强、细化、二值化、特征提取及模式识别才能获得矢量数据.人们通常将多色地图分色后逐个元素(如等高线地貌、水系、道路网、地物、符号与注记等)加以识别和提取.如果将数字影像矢量化，则需要事先做好重采样、图象处理、影像匹配和影像理解等过程，才能将影像上的语义和非语义信息提取出来，并形成矢量形式的数据.

数据压缩(data reduction or data compression) 处理数据的一种方法.是指减少数据冗余或对给定数据减少其存储空间的一种数据处理技术.其方法主要有：

1.削减.删去不必要的数据，被削减数据可用外延法或内插法补上.

2.参数抽出.即只保存特征数据和参数.

3.等间隔采样.即对连续输入的数据按等时间(或等增量、等距离)间隔采样.

4.编码变换.即把给出的数据变换成简化的代码予以存储.

5.函数应用.只保留尽可能少的特征点，被删去的数据点可采用某种函数进行恢复.

曲线光滑化(line smoothing) 地图上各种曲线自动绘制的方法.是根据给定的离散点列，用插值法或曲线拟合法建立符合实际要求的连续光滑曲线的函数，并按该函数计算加密点列和完成光滑连结的过程.在机助地图制图中，一般要求采用曲线插值法建立曲线函数，即根据给定离散点分段建立数学函数(通常用代数多项式或样条函数)，使函数图形通过给定离散点，并保持一阶或二阶导数连续.

触觉地图(tactual map) 地图的种类之一.它是一种提供盲人或低视力人专用的特制地图.用塑料制成或特殊的发泡油墨印制.使用时用手指触摸图上用凹凸线条或点构成的代表地区轮廓、主要地物形状和走向的简略地图，并附有盲文说明.有的也加印一二个简单颜色，以便视力正常的人(如盲校教师)参考.这种地图也有不同比例尺和区域，如街道图、市区图，直至很小比例尺的世界图.1991 年，中国第一次研制和出版了《中国触觉地图集》.

地图投影(map projection) 地图制图中一个重要的理论和方法.即在平面上表示地球面经纬线的理论和方法.依据一定的数学法则，把地球椭球(或球)表面经纬线网转化为平面上相应的经纬线网.或以一般数学解析式将二者的关系表达为

$$x=f_1(\varphi,\lambda),\quad y=f_2(\varphi,\lambda)$$

(φ,λ 是地球椭球(或球)表面点的地理坐标值，x,y 是平面上相应点的直角坐标值，函数 f_1,f_2 在一定域内必须是单值、有限而连续的).这样的理论和方法称为地图投影.

因为无论把地球作为椭球或球，在数学上都是不可展开的曲面，而地球是一个完整的曲面，在地图投影中把曲面描写到圆锥、圆柱或平面上时必然包含有各种变形.这种变形可以分为角度、长度和面积的变形.而一个地图投影只能保持某一方面量度的正确，而在另一些方面含有一定的误差，因此从变形的性质来划分，地图投影可以分为等角投影(或称保角或正形投影，是指一点上两微分线段构成的角度在投影后保持不变)、等面积投影(一定面积在投影后保持大小不变)、任意投影(所有既不等角，也不等面积的投影).在任意投影中，有一类通过一共同点的大圆族(例如经线或过球体上某一点的大圆系)保持长度不变的，称为等距离投影.常用投影又因采用的中间投影面的不同可分为圆锥投影、圆柱投影和方位投影(平面).也有些投影，没有中间投影面，而是依据一定的数学条件建立经纬线的平面表象的，可统称为条件投影.

等角投影(conformal projection) 见“地图投影”.

等面积投影(equal-area projection) 见“地图投影”.

任意投影(arbitrary projection) 见“地图投影”.

等距离投影(equidistant projection) 见“地图投影”.

条件投影(conditioned projection) 见“地图投影”.

地图集(atlas) 多幅地图的汇集.这些地图有协调一致的主题内容，按统一的设计原则和编制体例及规定的比例尺和分幅系统编制而成，其装帧形式有装订成册的或散页函装的.小型而内容较简单、图幅数不甚多的，也称之为地图册.地图集可按各种

标志划分，如按区域可分为世界地图集、洲地图集、国家地图集或区域（省、市、自治区、县、大河流域等）地图集；按内容可分为普通地图集、专题地图集、综合地图集，专题地图集更可细分成许多主题，如人口地图集、经济地图集、农业地图集等；按用途可分为教学地图集、旅游地图集等.

梅卡托投影（Mercator projection） 地图投影的种类之一. 它是用于航海与航空的地图投影. 由德国地理及地图学家梅卡托（Mercator，N.）发明. 它的原理是：以圆柱面切于赤道或割于南北两同名纬线上，按等角（即正形）条件把地球上的经纬线投影到圆柱面上，然后沿一母线把圆柱展开成平面，这样，经纬线描写成两组互相正交的直线，成为梅卡托投影的地图网格. 在这种地图上，航海所遵循的等角航线（亦称斜航线，loxodrome）成为连结起终两点的直线，便于航行作业，迄今仍广泛使用.

球心投影（gnomonic projection） 地图投影的种类之一. 它是以球心为发射中心的一种地图投影. 它将地球视为球体，一平面切该球面于一点（称此切点为投影中心）. 设想球心上有一发光点，可将球面上的经纬线投影到切平面上，构成球心投影网格，则称此种投影为球心投影. 在这种投影下，地面上任两点间的大圆弧（称为大圆航线或大环航线）表现为直线. 可用于航海、航空的航路设计，非常方便. 但这种投影只能描绘半球以内的范围，故视需要可将投影中心设置在不同位置，以获取地面不同区域的投影.

朗伯等角圆锥投影（Lambert conformal conic projection） 地图投影的种类之一. 它是一种等角（正形）圆锥投影. 由德国数学家朗伯（Lambert，J. H.）提出. 用一圆锥与地球相切于一条纬线或相割于两条纬线上，用等角（正形、保角）条件将经纬线投影到圆锥面上，然后沿圆锥一母线展成平面，则经线成为通过圆锥顶点的扇形等间隔直线，其夹角与经差呈正比，纬线为同心圆圆弧，称此投影为朗伯等角圆锥投影. 该投影中经纬线形状与地表实际情况接近，适用于中纬度地区的中、小比例尺地图. 中国 1：1000000地图及许多省、区图均采用该投影作为数学基础.

投影变换（projection transformation） 从一种地图投影变换成另一种地图投影的理论和方法. 各种地图的数学基础（即所采用的地图投影）往往并不相同，当运用一种地图作为资料来编制以另一种投影为数学基础的新地图时，就会出现投影变换的问题. 投影变换是两平面间的点的变换，也就是$(x,y)\to(X,Y)$，即建立两者之间的函数关系：

$$X = f_1(x,y);\quad Y = f_2(x,y).$$

实际作业中，有时不一定能建立直接的关系，而运用其他方法，如 $x,y\to\varphi,\lambda\to X,Y$，即由 x,y 求解一定点的经纬度 φ,λ，再由 φ,λ 用新投影的公式求得 X，Y. 还有其他各种数值变换法，如

$$X = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + \cdots,$$
$$Y = b_{00} + b_{10}x + b_{01}y + b_{20}x^2 + b_{11}xy + \cdots,$$

是由两个面上若干已知的对应点求解系数 a_{ij}，b_{ij}. 由此可求出其他点(x,y)变换后的坐标(X,Y). 投影变换在计算机制图与遥感制图中有重要用处.

心像地图（mental map） 亦称意境地图. 地图的种类之一. 在理论地图学中，一种非常规制作的地图. 它是由人对环境的认识在记忆中构成印象，依据印象而描绘出的示意性概略地图. 这种地图的正确程度与人的文化水平、对地理环境接触与认识程度、对所描绘地区居住时间长短等因素有关，是测验人的智力的一种图解方式.

意境地图（mental map） 即"心像地图".

拓扑地图（topological map） 地图的种类之一. 是指地图学中一种统计地图，一种保持点与线相对位置关系正确而不一定保持图形形状与面积、距离、方向正确的抽象地图. 例如一个国家轮廓可以简化成多边形或纵横直线联成的图形，它与周边国家接壤的相对关系保持正确，但面积可以按某种统计指标来反映其大小.

互补色地图（anaglyphic map） 地图的种类之一. 按互补色原理制作的地图. 利用互补的红青两色，将两个位置拍摄的同一地区影像按视差叠印在一起，然后双眼分别通过用红青两色制成的眼镜，可以看到该地区的立体模型，称之为互补色地图. 如果对等高线地形图也同样处理，将等高线及地物按其高低构成不同的视差，分别制成两种颜色的地形图，按规定视差叠印，用同样方法可看出立体地形形象.

比较地图学（comparative cartography） 理论地图学的一个分支. 根据人类善于利用比较方法认识事物差异的能力，对地图学的一些基本问题的研究作方法学的探讨的科学.

元地图学（metacartography） 一种超地图学的理论. 从哲学的角度对地图学的最基本问题，如地图传输、地图符号系统、地图语言、地图认知过程的规律进行本质研究的学科. 关于这方面较系统研究的著作，迄今仅有苏联格鲁吉亚共和国学者阿斯兰卡斯克维里（Aslanikaskvili）发表的专著《Метакартография》. 该书论述了地图学的理论，试图把地图学各个部分统一到一个单一的逻辑方法学系统.

变比例尺投影（varioscale projection） 地图投影的种类之一. 它能在图面上造成两个或更多不同比例尺图形部分的地图投影. 是利用不同地图投影之间变换时变形变化的规律，使地图的某些重要部分具有较大的比例尺（放大）而压缩次要的部分（缩

小)的投影方法.对于城市或某些旅游地图特别适用.也可以在整个图面上突出一个以上的局部地区,称为多焦点投影.现代的方法已发展到设定放大倍数和一定的范围,以及变比例尺局部的不同形状,使得更适应于实际的要求.

多焦点投影(multi-focuses projection) 见“变比例尺投影”.

通用横梅卡托投影(universal transverse Mercator projection) 简称 UTM 投影.地图投影的种类之一.它是横轴等角圆柱投影.其结构相当于高斯-克吕格投影,但其中央经线的长度比后者缩短了一个常数,这样能使投影区域的相对变形比后者有所改善.该投影被许多西方国家作为地形图的数学基础.

丝网印刷(silk-screen printing) 亦称网版印刷.印刷的种类之一.是一种用丝网作为印刷基础的印刷技术,因其印版的基材而得名.它是将丝网紧绷于坚固网框上,用适当的材料和方法,堵塞非图文区的网孔而成模版,然后注油墨于框内的网版上,在刮墨刀的刮动压力作用下,油墨即由图文区的网孔漏印到承印物上,即得所需之印迹.丝网印刷具有印刷墨层厚(6～100μm)、适印的油墨和承印物广、操作容易及投资小等特点,故在电子、印染、塑料、陶瓷、标牌及印刷等行业中,广泛用于印刷纸、木、布、塑料、金属、玻璃及陶器等物品,在包装印刷中占有重要的位置.

网版印刷(silk-screen printing) 即“丝网印刷”.

四色印刷(four color printing) 印刷的种类之一.它是用黄(Y)、品红(M)、青(C)和黑(BK)四种颜色来进行彩色印刷的一种方法.理论上,四色印刷可以获得成千上万种颜色,以 Y,M,C 和 BK 四种色组合,而不重复.单色——Y,M,C 和 BK 四种;双色——Y+M,Y+C,Y+BK,M+C,M+BK,C+BK 六种;三色——Y+M+C,M+C+BK,Y+C+BK,Y+M+BK 四种;四色——Y+M+C+BK 一种,共计十五种颜色.彩色印刷时,是利用“网点”的大小来表示颜色的深浅,在理论上,“网点”可分为 22 个等级,根据排列组合公式$C_m^n \cdot g^n$来计算,可获得 $C_4^1 \cdot 22^1 + C_4^2 \cdot 22^2 + C_4^3 \cdot 22^3 + C_4^4 \cdot 22^4 = 279840$(种)颜色.

实际上,由于印刷工艺过程中“网点”的形变误差以及视觉辨认阈限的限制,四色印刷所能够获得的彩色比理论上要少得多.

减色印刷(reducing color printing) 印刷的种类之一.是用减色法原理进行彩色印刷的一种方法.光不仅是人类生存的基础,而且是人的视觉感受到颜色的先决条件.人们能够感受到的色光为红、橙、黄、绿、青、蓝、紫,称可见光谱,其中任意两种或两种以上色光相混合,呈现的第三种色光的亮度,是各自色光亮度之和,因而称色光的混合为加色法.各种颜料、染料或油墨对于白光,都具有反射一定波长的色光和吸收其余光的物理特性,因而某种颜色的油墨,呈现出反射某一段波长的颜色.两种色料相混合而呈现的第三种颜色明度降低,若再加入一种色料,则明度会继续降低,因此称色料的混合为减色法.彩色印刷是利用不同颜色的油墨,将大小不同的“网点”叠印在白纸上,也是属于色料的混合,故称为减色印刷.

平版印刷(lithography) 印刷的种类之一.它是用图文和空白两部分几乎在同一平面上的印版来进行印刷的一种印刷方法.印刷时,根据油和水难以相融原理,图文部分疏水亲油,空白部分疏油亲水,将印版上的图文经橡皮布转移到承印物上,属于一种间接印刷方式,俗称为胶印.平版印刷具有制版容易、幅面大、生产周期短和成本低等特点.因使用橡皮布,印刷压力小,印刷品网点清晰、层次丰富、色调柔和,广泛应用于彩画、地图和装潢印刷等,成为一种十分重要的印刷方法.随着激光照排技术的发展,昔日以凸版印刷为主的报纸和书刊也正逐步被平印所淘汰.但由于印刷时要用药水,容易出现油墨乳化、印迹浅淡和干燥慢等弊病,目前正在开展的无水平印研究已获得初步成功,将是平版印刷的一次重大变革.

测量误差及平差

最小二乘平差(least squares adjustment) 数学的一个分支学科.它是按最小二乘原理对观测数据进行处理,以求出未知参数最佳估值的一个学科.

最小二乘平差的基本要素是函数模型和随机模型.根据函数模型的不同,通常将最小二乘平差分为条件平差、附有未知数的条件平差、参数平差和附有限制条件的参数平差.此后,不少文献中又加入了另一个函数模型,即附有限制条件的条件平差模型.20世纪 80 年代於宗俦首先提出该模型可以概括上述各种模型,即其他模型都可以看做是该模型的特例,因此称它为概括平差模型,并由它导出了适用于各种平差方法的通用公式.应用概括平差模型及其通用公式进行理论推导,其结论将适用于所有平差方法,这就为测量平差的某些理论研究提供了一个更为有效的途径.此外,根据随机模型的不同,又将最小二乘平差分为独立观测平差和相关观测平差,前者则是后者的特例.

在测量平差的计算工作中,法方程组的计算占着很大的比例.在电子计算机出现之前,一个高阶法

方程组的计算，不仅花费时间多，而且难以保证计算的精度. 不少测量学者为此相继提出了一些“化整为零”的计算方法，如分组平差、分区平差、序贯平差等. 即使在有了大型快速计算机的今天，这些方法仍然能起到节省机时、解决计算机内存容量不敷应用以及能实时提供平差成果等作用.

20 世纪 60 年代以后，测量平差理论又有了一些新的进展，例如：

1. 提出了参数平差中系数阵列不满秩时的秩亏网平差(亦称自由网平差).

2. 将观测向量满秩权逆阵扩展到奇异阵的情况.

3. 将通信和自动控制领域中的数据处理方法应用于测量数据处理的最小二乘滤波、推估和配置的方法，即所谓近代平差法.

4. 提出削弱系统误差影响、粗差检验和定位的理论和方法，提出测量平差系统的可靠性研究理论，李德仁由两个多维备选假设出发，提出了平差系统的可区分性和可靠性理论，拓宽了国外的研究成果.

5. 提出了方差-协方差分量验后估计的理论和方法.

上述新的理论和方法的提出，使测量平差的概念更加广义化，理论研究更加深入了.

概括平差模型(universal adjustment model) 见“最小二乘平差”.

测量误差(measuring error) 测量平差中的一个术语. 它是指被观测量在一定度量下真值与观测值之差. 误差产生的原因有仪器因素、人的感官因素以及观测时外界自然条件的影响等，因此任何测量总不可避免地伴随着误差. 被观测量的真值与观测值之差称为真误差. 按测量误差的性质可分为粗差、系统误差、偶然误差. 在测量平差中，主要是研究和处理有关偶然误差的问题. 但从 20 世纪 60 年代以来，在测量平差中越来越多地研究和讨论有关系统误差补偿和改正以及粗差定位和剔除的有关理论和方法.

真误差(true error) 见“测量误差”.

偶然误差(random error) 亦称随机误差. 测量误差的种类之一. 在相同观测条件下进行的一系列观测中，从表面现象看，其数值和符号没有规律性，但按其数值和符号作分类统计，却呈现出一定规律性. 在一般情况下，偶然误差的出现服从正态分布. 它具有如下统计规律性：

1. 绝对值相同、符号相反的偶然误差出现的概率相等.

2. 绝对值大的比绝对值小的偶然误差出现的概率小.

3. 在一定的观测条件下，误差的绝对值有一定范围，超过这个范围的误差出现的概率极小，通常认为它是不可能出现的.

4. 偶然误差的数学期望为零.

以上这些性质是处理偶然误差和衡量精度的基础.

随机误差(stochastic error) 即“偶然误差”.

系统误差(systematic error) 测量误差的种类之一. 在相同观测条件下进行的一系列观测中，数值和符号或保持不变或按一定规律变化着的误差. 系统误差对观测成果的影响具有积累性，故对成果质量影响显著. 但它可以通过对仪器和自然条件的影响进行校验和测定，以及选择适当的测量方法和程序等予以消除和改正. 尽管如此，在平差前还是无法保证能完全消除系统误差的影响. 近年来又提出了在平差的函数模型中加入附加参数的办法，以便通过平差来消除系统误差对平差结果的影响(参见“自检校平差”).

粗差(gross error) 测量误差的种类之一. 一般是指绝对值大于 3 倍中误差的观测误差，包括内外业中因疏忽大意而造成的差错在内. 粗差的存在将大大影响平差结果的可靠性，甚至导致完全错误的结果. 传统上是在进行测量外业和内业的过程中，通过一系列的措施，例如采用适当的观测程序，进行可供检核的重复观测，增加多余观测，利用几何条件的闭合差大小加以限制等，及时发现并限制粗差，发现后予以剔除. 尽管如此，仍然难免会有残存的小粗差存在. 因此，从 20 世纪 60 年代末起，又相继发展了一些处理粗差的理论和方法，其中有以统计假设检验为基础的粗差检验法，如荷兰巴尔达(Baarda, W.)教授提出的数据探测法等，此外还有以稳健估计为基础的选权迭代法等. 从而在平差计算中实现了粗差自动剔除的目的(参见“粗差检测”、“数据探测法”、“选权迭代法”).

常差(constant error) 测量误差的种类之一. 它是指数值和正负号均保持不变的误差. 是系统误差中的一种.

人仪差(personal and instrumental equation) 系统误差中的一种. 它是指天文观测中的一种误差. 在天文经度(或时间)测定中，由于观测者的生理和心理的特点，使观测成果中带有一种系统性的偏差. 这种偏差因人而异，同一人在不同时间、条件、状态下也会不同，故称人差. 由于所用仪器的光学和机械结构的特性，在观测结果中引起的系统性偏差，称为仪器差. 这两种偏差常常混合在一起，统称为人仪差. 人仪差可能是正值，也可能是负值，但它是系统性的误差，需从观测结果中消除. 在天文作业的前后，一般都要在经过精密测定的天文基本点上测定人仪差，以便在观测结果中加以改正. 人仪差不仅存

在于天文经度测定中，也存在于天文纬度和天文方位角测定中.

平均误差(average error) 衡量观测精度的指标之一，它是真误差绝对值之和的平均值. 设在一定观测条件下，出现的一组独立的真误差为 $\Delta_1,\Delta_2,\cdots,\Delta_n$，则平均误差定义为

$$\theta=\lim_{n\to\infty}\frac{|\Delta_1|+|\Delta_2|+\cdots+|\Delta_n|}{n}=\lim_{n\to\infty}\frac{[|\Delta|]}{n}.$$

若 n 为有限时，则平均误差的估值为

$$\hat{\theta}=\pm\frac{[|\Delta|]}{n}.$$

从理论上证明，平均误差与中误差的关系是

$$\theta=0.7979\sigma.$$

式中 σ 为中误差(参见“方差”).

或然误差(probable error) 衡量观测精度的指标之一. 它是指在一定条件下，其绝对值比它大的误差和绝对值比它小的误差出现的概率相等，即

$$P(|\Delta|\leqslant\rho)=\frac{1}{2},$$

式中 ρ 表示或然误差. 从理论上证明，或然误差与中误差的关系为 $\rho=0.6745\sigma$.

相对误差(relative error) 衡量相对精度的一个指标. 它是某量误差的绝对值与该量本身大小之比. 通常是用分子为 1 的分数表示，相对误差常用来表征边长观测值的精度. 设测量某边长 l 时，其中误差为 σ，则其相对中误差为 $1/N$ $(N=l/\sigma)$. 若用该边长的极限误差，则称为相对极限误差. 相对误差是一个无名数.

限差(tolerance) 亦称极限误差或容许误差. 测量平差中的一个术语. 它是指在一定条件下偶然误差的绝对值不应超过的限值. 它可用来检核观测质量，决定成果取舍. 按照概率可以证明，在一系列观测中误差的绝对值大于中误差 σ 出现的机会约为 1/3；大于两倍 σ 出现的机会约为 1/22；大于三倍 σ 出现的机会约为 1/370. 因此，在测量工作中，通常规定以 2σ 或 3σ 作为误差的限差，误差绝对值超过限差时，该观测值即予以删去或重测.

极限误差(limit error) 即“限差”.

容许误差(allowable error) 即“限差”.

方差(variance) 测量平差中最常用的精度概念之一. 它是衡量观测精度的一种最常用指标，是描述随机变量离散度的特征数字. 设在一定的观测条件下，出现一组独立的真误差 $\Delta_1,\Delta_2,\cdots,\Delta_n$，则定义方差为

$$\sigma^2=E(\Delta^2)=\lim_{n\to\infty}\frac{\Delta_1^2+\Delta_2^2+\cdots+\Delta_n^2}{n}=\lim_{n\to\infty}\frac{[\Delta^2]}{n}.$$

方差愈小，表示精度愈高. 当误差的个数 n 为有限时，求得的是方差的估值

$$\hat{\sigma}^2=\frac{[\Delta^2]}{n}.$$

方差的平方根 σ 称为标准差，在测量学中习惯上称为中误差或均方误差.

中误差(mean square error) 见“方差”.

均方误差(mean square error) 见“方差”.

标准差(standard deviation) 见“方差”.

协方差(covariance) 测量平差中最常用的精度概念之一. 它是描述随机变量之间(统计)相关程度的特征数字. 设有随机变量 X 和 Y，在一定的观测条件下，出现两组相互独立的真误差分别为 $\Delta_{x_1},\Delta_{x_2},\cdots,\Delta_{x_n}$ 和 $\Delta_{y_1},\Delta_{y_2},\cdots,\Delta_{y_n}$，则 X 和 Y 之间的协方差定义为

$$\begin{aligned}\sigma_{xy}&=E(\Delta_x\Delta_y)\\&=\lim_{n\to\infty}\frac{\Delta_{x_1}\Delta_{y_1}+\Delta_{x_2}\Delta_{y_2}+\cdots+\Delta_{x_n}\Delta_{y_n}}{n}\\&=\lim_{n\to\infty}\frac{[\Delta_x\Delta_y]}{n}.\end{aligned}$$

协方差愈大，表示 X 和 Y 之间的相关程度愈强，反之，表示两者之间的相关程度愈弱. 当协方差 $\sigma_{xy}=0$ 时，则表示 X 与 Y 之间互不相关. 当误差对数 n 为有限时，求得的是协方差的估值

$$\hat{\sigma}_{xy}=\frac{[\Delta_x\Delta_y]}{n}.$$

权(weight) 测量平差中最常用的精度概念之一. 它是用以表征精度高低的一种相对指标. 设有观测值 L_i，它的方差为 $\sigma_i^2(i=1,2,\cdots,n)$，则定义

$$p_i=\frac{\sigma_0^2}{\sigma_i^2},$$

其中 σ_0^2 是任意选定的某个比例常数，p_i 即表示 L_i 的权. 权有如下性质：

1. 权与方差成反比，权愈大，表示精度愈高，权愈小，表示精度愈低.

2. 一组观测值的权，其数值大小随所选比例常数 σ_0^2 的不同而改变. 若选定另一比例常数 σ'^2_0，则得另一组相应的权 $p'_i(i=1,2,\cdots,n)$，但不论选定何值，各组权之间的比例关系不变，即

$$p_1:p_2:\cdots:p_n=p'_1:p'_2:\cdots:p'_n.$$

3. 权是表征观测精度高低的比例数，它是一种相对指标. 因此，在同一个平差问题中，为了反映精度高低的比例关系，只能选用同一个数值而不能同时选用不同的数值作为定权时的比例常数.

单位权(unit weight) 测量平差中最常用的精度概念之一. 它是指数值等于 1 的权(参见“单位权方差”).

单位权方差(variance of unit weight) 测量平差中最常用的精度概念之一. 它是指权等于 1 所对应的方差. 观测值 L_i 的权定义为

$$p_i=\frac{\sigma_0^2}{\sigma_i^2},$$

式中 σ_0^2 是定权时所选用的比例常数．若某个观测值 L_i 的方差 σ_i^2 等于比例常数 σ_0^2 时，则它的权必然等于1，可见 σ_0^2 就是权等于1的方差，因此习惯上将定权的比例常数 σ_0^2 称为单位权方差，同时将数值为1的权称为单位权，又将权等于1所对应的观测值称为单位权观测值．

单位权观测值(observation of unit weight)　测量平差中的一个术语．它是指权等于1所对应的观测值(参见“单位权方差”)．

方差-协方差(矩)阵(variance-covariance matrix)　亦称积差阵，简称协方差阵．随机向量精度的表示方法之一．它是由一个随机向量中各元素的方差和两两元素间的协方差按一定排列顺序构成的矩阵．例如，有一个 $(n\times1)$ 维的观测向量 $L=[L_1\ L_2\ \cdots\ L_n]^{\mathrm{T}}$，它的方差-协方差阵可表示为

$$D_L=\begin{bmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n}\\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n}\\ \vdots & \vdots & & \vdots\\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2\end{bmatrix},$$

式中位于主对角线上的 σ_i^2 是观测值 $L_i(i=1,2,\cdots,n)$ 的方差，称为主对角元素，所有的非主对角元素是两两观测值之间的协方差．在第 i 行中，除了包含 L_i 的方差 σ_i^2 之外，位于第 i 行第 j 列 $(j\neq i)$ 的元素则相应为 L_i 关于 L_j 的协方差．因 $\sigma_{ij}=\sigma_{ji}$，故方差-协方差阵是一对称方阵．

方差-协方差阵全面表征了一个随机向量中各元素的精度和相互间的相关程度．当协方差 $\sigma_{ij}(i\neq j)$ 全部等于零时，表示各观测值之间均互不相关，方差-协方差阵即变成对角阵．

积差阵(variance-covariance matrix)　即“方差-协方差矩阵”．

协因数(cofactor)　测量平差中有关精度的一个重要概念．它是指观测值的权倒数或两个观测值之间的相关权倒数．设某观测值 $L_i(i=1,2,\cdots,n)$ 的方差为 σ_i^2，L_i 与 L_j 之间的协方差为 $\sigma_{ij}(i\neq j,i,j=1,2,\cdots,n)$，则定义：

$$Q_{ii}=\frac{\sigma_i^2}{\sigma_0^2},\quad Q_{ij}=\frac{\sigma_{ij}}{\sigma_0^2},$$

Q_{ii} 称为 L_i 的协因数，Q_{ij} 称为 L_i 与 L_j 之间的互协因数．根据权的定义知，L_i 的权为

$$p_i=\frac{\sigma_0^2}{\sigma_i^2},$$

故知 L_i 的协因数 Q_{ii} 实际上就是 L_i 的权倒数．类似地称 Q_{ij} 为 L_i 和 L_j 的相关权倒数．

协因数(矩)阵(cofactor matrix)　随机向量精度的表示方法之一．它是由一个随机向量中各元素的协因数和两两元素之间的互协因数按一定顺序排列而成的矩阵．例如，设有 $(n\times1)$ 维的观测向量 $L=[L_1\ L_2\cdots\ L_n]^{\mathrm{T}}$，它的协因数阵可表示为

$$Q_L=\begin{bmatrix}Q_{11} & Q_{12} & \cdots & Q_{1n}\\ Q_{21} & Q_{22} & \cdots & Q_{2n}\\ \vdots & \vdots & & \vdots\\ Q_{n1} & Q_{n2} & \cdots & Q_{nn}\end{bmatrix},$$

式中主对角元素 $Q_{ii}(i=1,2,\cdots,n)$ 是观测值 L_i 的协因数，非主对角元素 $Q_{ij}(i\neq j,i,j=1,2,\cdots,n)$ 是观测值 L_i 与 L_j 的互协因数．因为

$$Q_{ii}=\frac{\sigma_i^2}{\sigma_0^2}=\frac{1}{p_i},\quad Q_{ij}=\frac{\sigma_{ij}}{\sigma_0^2},$$

Q_{ii} 就是第 i 个观测值的权倒数，相应地称 Q_{ij} 为 L_i 与 $L_j(i\neq j)$ 的相关权倒数，因此协因数阵又称权逆阵．当观测值互不相关时，非主对角元素全部等于零，协因数阵即成为对角阵．

权逆阵(inverse weight matrix)　见“协因数(矩)阵”．

权阵(weight matrix)　随机向量精度的表示方法之一．它是指协因数阵的逆阵．设有 $(n\times1)$ 维随机向量，其协因数阵为 Q，权阵为 P，则权阵定义为

$$P=Q^{-1}=\begin{bmatrix}Q_{11} & Q_{12} & \cdots & Q_{1n}\\ Q_{21} & Q_{22} & \cdots & Q_{2n}\\ \vdots & \vdots & & \vdots\\ Q_{n1} & Q_{n2} & \cdots & Q_{nn}\end{bmatrix}^{-1}=\begin{bmatrix}P_{11} & P_{12} & \cdots & P_{1n}\\ P_{21} & P_{22} & \cdots & P_{2n}\\ \vdots & \vdots & & \vdots\\ P_{n1} & P_{n2} & \cdots & P_{nn}\end{bmatrix}.$$

协因数阵和权阵互为逆阵，即 $P=Q^{-1}$ 或 $Q=P^{-1}$，因此 $PQ=QP=I$．需要指出的是，协因数阵是权逆阵，该矩阵中的各个元素都具有权倒数的意义．例如，它的主对角元素 Q_{ii} 就是第 i 个随机变量的权的倒数．但作为一个随机向量的权阵，它的各个元素不一定就是该向量中各个随机变量的权．例如，权阵的主对角元素 P_{ii} 并非是第 i 个随机变量的权，因为 $P_{ii}\neq Q_{ii}^{-1}$．只有当各随机变量互不相关，其协因数阵成为对角阵时，再由它求逆而得到的权阵，其各主对角元素才分别是各随机变量的权．

精度估计(precision estimation)　亦称精度评定．测量平差中常用于精度衡量的一个术语．通过对观测数据的处理，求得各种量的精度．测量平差的两个主要任务：一是确定各种量的平差值(即估值)，另一是评定精度．精度评定主要包括估计观测值、平差值(观测量平差值、参数平差值等)以及它们的函数的精度．用以估计精度的主要指标是方差、协方差以及表征相对精度的协因数和权等．协方差传播律和

协因数传播律则是评定观测值函数或平差值函数精度时所要用到的重要计算公式.

精度评定(precision evaluation)　即"精度估计".

方差-协方差传播律(variance-covariance propagation law)　简称协方差传播律.测量平差中用于精度衡量的一个重要公式.它是根据随机变量的方差和协方差求其函数的方差和协方差的计算公式.评定精度是平差的主要任务之一,因此协方差传播律在测量平差中起着重要的作用.设有

$$Z = K_1X + K_2Y + Z^0, \quad (1)$$

$$W = K_3X + K_4Y + W^0, \quad (2)$$

式中 Z 和 W 都是随机向量 X 和 Y 的线性函数.已知 X 和 Y 的协方差阵分别为 D_X 和 D_Y,X 关于 Y 和 Y 关于 X 的互协方差阵为 D_{XY} 和 D_{YX},K_1,K_2,K_3,K_4 都是已知的系数阵,Z^0 和 W^0 是函数中的常数项(非随机量).若要求函数 Z 的协方差阵 D_Z,则可略去(1)式中对求协方差无关的 Z^0 项,写出(1)式和它的转置式

$$Z = K_1X + K_2Y,$$

$$Z^T = X^TK_1^T + Y^TK_2^T,$$

按代数的乘法规则作乘法运算,并将各乘积项中两随机向量的乘积(如 ZZ^T,XX^T,XY^T,YX^T,YY^T 等),相应地写成 D_Z,D_X,D_{XY},D_{YX},D_Y,则得

$$D_Z = K_1D_XK_1^T + K_1D_{XY}K_2^T + K_2D_{YX}K_1^T + K_2D_YK_2^T. \quad (3)$$

若要求函数 Z 关于 W 的互协方差阵 D_{ZW},则可略去(1)和(2)式中的 Z^0 和 W^0 项,写出(1)式,并取(2)式的转置,按上述同样的方法,即得

$$D_{ZW} = K_1D_XK_3^T + K_1D_{XY}K_4^T + K_2D_{YX}K_3^T + K_2D_YK_4^T. \quad (4)$$

(3)和(4)式统称为协方差传播律公式.当函数为非线性形式时,只需先求该函数的全微分,并将对各变量的偏导数值组成已知的系数阵,然后即可应用协方差传播律的公式计算所需要的协方差阵或互协方差阵.

协因数传播律(cofactor propagation law)　测量平差中用于精度衡量的一个重要公式.它是根据随机变量的协因数和互协因数求其函数的协因数和互协因数的计算公式.由于协因数阵与协方差阵之间仅差一个因子 σ_0^2,因此只要将协方差传播律有关公式中的协方差阵和互协方差阵换成相应的协因数阵和互协因数阵,例如将 D_X 换成 Q_X,将 D_{XY} 换成 Q_{XY} 等,就得到了协因数传播律的公式(参见"方差-协方差传播律").

最小二乘准则(least squares criterion)　进行最小二乘平差计算的一个基本原则.它是求解不定线性方程组的一个附加条件.在任何平差计算中,所列出的方程式的个数,总是少于方程中所包含的未知量的个数,因此其解不惟一.在最小二乘准则下求解,可以得到一组惟一解.若在平差中,只有观测值为随机量时,最小二乘准则为

$$V^TPV = \min;$$

如果不仅仅观测值为随机量,而且参数也是随机量时,则最小二乘准则为

$$V^TPV + \hat{x}^TP_x\hat{x} = \min,$$

并称它为广义最小二乘准则.V 是观测向量的改正数向量,P 是观测向量的权阵,$\hat{x}$ 是参数向量的改正数向量,P_x 是参数向量的权阵.

广义最小二乘准则(generalized least squares criterion)　见"最小二乘准则".

必要观测(necessary observation)　测量平差中的一个术语.它是指为确定某个几何或物理模型所必须进行的最少个数的观测.必要观测的个数、类型和要求是:

1.必要观测的个数随所选定的模型而定.例如,确定一平面三角形的形状(这是一种几何模型),其必要观测数为2,它们可以是其中任意的两个角度;又如,确定其形状和大小(这是另一种几何模型),则其必要观测数为3,它们可以是其中任意的两角一边,或一角两边,或三边.必要观测数常用 t 表示.

2.必要观测的类型应与模型相适应.例如,三角形的形状可以通过其中的任意两个角度来确定,但要确定其大小,则在三个必要观测中至少要有一个边长,仅仅三个内角是无法确定其大小的.

3.在 t 个必要观测中,其中任一个都不可能由其余 $(t-1)$ 个所导出,即在这 t 个必要观测之间不存在任何函数关系,它们都是函数独立的量.例如,三角形的三个内角之间存在着一个几何关系,故其中只有两个内角可作为必要观测.

多余观测(redundant observation)　测量平差计算中的一个术语.它是指在一组观测值中,除了能惟一确定某个几何或物理模型的 t 个必要观测之外的其余观测值.设在一系列观测中共有 n 个观测值,其中除了 t 个必要观测之外,其余的 $(n-t)$ 个即为多余观测 $(n>t)$.多余观测的个数常用 r 表示,即 $r=n-t$.多余观测数又称平差的自由度.在 n 个观测中,除了 t 个必要观测之间不存在任何函数关系外,随后每多增加一个多余观测或未知数(非观测量),就会在它们之间多产生一个函数关系式,在测量平差中称之为条件方程(参见"条件方程").条件方程是进行平差计算的依据,因此没有多余观测,也就不存在平差问题.

通过多余观测,不仅可以检核观测值中是否含有超过允许范围的误差甚至错误,而且在平差后可

以提高计算成果的精度.

函数模型(functional model) 测量平差的基本要素之一. 它是描述一些量之间所存在的几何或物理的函数关系,是描述观测值期望的数学模型. 参与平差计算的,除了观测量之外,有时还需要加入一些非观测量作为参数(或称未知数). 当所选参数之间不存在确定的函数关系时,称为(函数)独立参数;否则,称为不独立参数. 函数关系表达式有线性的和非线性的. 在平差计算中,需将非线性函数化为线性形式. 平差时是否加入参数、参数是否独立以及参数个数的不同,就构成了不同的函数模型,其中具有最一般形式的线性函数模型为

$$\underset{c,n}{A}\underset{n,1}{V}+\underset{c,u}{B}\underset{u,1}{\hat{x}}-\underset{c,1}{f}=0, \tag{1}$$

$$\underset{s,u}{C}\underset{u,1}{\hat{x}}-\underset{s,1}{f_x}=0. \tag{2}$$

(1)式是附有参数的条件方程,(2)式是反映参数之间函数关系的限制条件方程(或称约束条件方程). A,B,C 是已知的系数阵,$\underset{n,1}{V}$ 和 $\underset{u,1}{\hat{x}}$ 分别是观测值和参数近似值 X^0 的改正数向量,$\underset{c,1}{f}$ 和 $\underset{s,1}{f_x}$ 是方程的常数项.

在最小二乘准则下解出上述函数模型中的 V 和 $\hat{x}$ 的方法,称为附有限制条件的条件平差法. 当参数独立,即 $C=0$ 时(这时(2)式不存在),称为附有未知数的条件平差法;当不加入参数,即 $B=0$ 和 $C=0$ 时,称为条件平差法;当 $A=-I$ (I 为单位阵)时,称为附有限制条件的参数平差法;当参数独立,即 $C=0$ 且 $A=-I$ 时,称为参数平差法. 可见后几种方法都是前一种方法的特例,因此将(1),(2)式中的函数模型称为概括平差模型或通用平差模型. 由该模型所导出的平差计算和精度评定公式称为通用公式,其他几种平差方法的公式均可由它简化而得.

在上述各平差方法中,观测值为随机量,参数为非随机量. 但在有些情况下,不仅观测值,而且参数也可能是随机量(参见"最小二乘滤波"、"最小二乘配置").

随机模型(stochastic model) 测量平差的基本要素之一. 它是用于描述随机变量精度以及它们之间随机性质的模型. 就观测向量 L 而言,表征其各元素的精度和彼此间统计相关程度的数字特征是方差和协方差,由它们所构成的方差-协方差阵 D_L (简称协方差阵)则是全面反映该向量统计性质的随机模型. 最小二乘平差就是按最小二乘准则,在给定函数模型和随机模型的情况下,由一组观测值求出未知参数的最佳估值,并估计其精度.

已知 $D_L=\sigma_0^2Q_L=\sigma_0^2P_L^{-1}$,式中 σ_0^2 是单位权方差,Q_L 和 P_L 分别是观测向量的协因数阵和权阵. 在平差计算中,当 Q_L 为单位阵时,表示观测值不相关且精度相等;当 Q_L 为对角阵时,表示观测值不相关且精度不等;当 Q_L 为非对角阵时,表示观测值相关且精度不等. 在前两种随机模型下的平差,称为独立观测平差,在后一种随机模型下的平差,称为相关观测平差.

模型误差(model error) 测量平差中的一个术语. 测量平差的两个基本要素为函数模型和随机模型,这两种模型统称为测量平差的数学模型. 模型误差就是指当数学模型不严格符合于它们所描述的客观实际所产生的误差. 为了根据一组观测数据估计未知参数,首先必须建立一个能反映观测数据与未知参数之间关系的数学模型,包括描述观测值期望的函数模型和描述观测值精度特性的随机模型. 但在实际问题中,无论是函数模型还是随机模型,要使它们严格符合客观实际,几乎是不可能的. 例如,在函数模型中未能反映出观测值中可能存在的系统误差或粗差,在随机模型中所给定的方差或协方差不够准确等,都会使两者存在一定的模型误差.

20 世纪 60 年代以后,相继发展和提出了一些消除模型误差影响的理论和方法. 例如,在平差函数模型中加入附加参数,以便在平差过程中消除系统误差的影响;通过数据探测、选权迭代等方法进行粗差检测,并实现粗差自动剔除(参见"自检校平差"、"粗差检测"、"数据探测法"、"选权迭代法"等);通过平差的验后信息,对方差、协方差分量进行估计,从而改善随机模型以达到提高平差结果精度的目的.

条件方程(condition equation) 平差函数模型中的基本方程之一. 它是用以描述某些几何量或物理量之间存在的函数关系式. 在测量平差中,将通过实测取得观测数据的量称为观测量,将未进行实测的量称为非观测量,习惯上将前者的平差值用符号 $\hat{L}$ 表示,后者的平差值用 $\hat{X}$ 表示. 它们之间存在的函数关系式常用的有如下几种形式:

$$F(\hat{L})=0, \tag{1}$$

$$F(\hat{L},\hat{X})=0, \tag{2}$$

$$F(\hat{X})-\hat{L}=0 \quad (\text{常写成 } \hat{L}=F(\hat{X})). \tag{3}$$

上述几种函数式统称为条件方程. 但为了便于区分起见,将不带未知数的条件方程(1)式仍称为条件方程,而将(2)式称为附有未知数的条件方程,将(3)式称为观测方程(参见"观测方程").

限制条件方程(constraint equation) 亦称约束条件方程. 平差函数模型中的基本方程之一. 它是描述未知数(或称参数)之间存在的函数关系式. 在平差计算中,除了观测量之外,为了某种目的,往往还要选定一些非观测量作为未知数参与平差. 未知数平差值常用符号 $\hat{X}$ 表示. 例如,在一平面三角形中,若将其中三个内角都选作未知数,其平差值分别为 $\hat{X}_1,\hat{X}_2,\hat{X}_3$,则在它们之间应存在这样的一个条

件方程：

$$\hat{X}_1 + \hat{X}_2 + \hat{X}_3 - 180° = 0.$$

这一条件方程的特点是：式中只包含未知数而无任何观测量. 这一类条件方程统称为限制条件方程，它的一般表达式为

$$\Phi(\hat{X}) = 0.$$

约束条件方程(constraint equation)　即"限制条件方程".

观测方程(observation equation)　平差函数模型中的基本方程之一. 它是一种特殊形式的条件方程，即在每个条件方程中仅仅包含一个观测量. 在参数平差和附有限制条件的参数平差中都要求将每一个观测量表达成所选参数的函数，其一般形式为

$$\hat{L} = F(\hat{X}),$$

式中

$$\hat{L}=L+V, \quad \hat{X}=X^0+\hat{x},$$

$\hat{L}$ 和 $\hat{X}$ 分别是观测量和参数的平差值，V 和 $\hat{x}$ 分别是观测值 L 和参数近似值 X^0 的改正数. 上述方程可能是线性的或非线性的，在平差计算前都要化为线性形式，线性化后的一般形式为

$$V = B\hat{x} - f.$$

上式称为误差方程. 其中 B 是已知的系数阵，f 是方程中的常数项.

误差方程(error equation)　测量平差中的一个术语. 它是指线性化后的观测方程(参见"观测方程").

平差值(adjusted value)　亦称估值或最或是值. 测量平差中的一个术语. 它是由平差计算所求得的各种结果的统称. 例如，有观测值的平差值、参数的(或未知数)平差值以及由它们所算得的某个函数的平差值等. 当前常用平差值和估值，而最或是值这一名词已较少采用.

估值(estimated value)　即"平差值".

最或是值(best possible value)　即"平差值".

联系数(correlate)　测量平差中的一个术语. 即数学中求条件极值的拉格朗日乘数. 以概括平差函数模型(附有限制条件的条件平差)为例，其线性化的条件方程为

$$\underset{c,n}{A}\,\underset{n,1}{V} + \underset{c,u}{B}\,\underset{u,1}{\hat{x}} - \underset{c,1}{f} = 0, \tag{1)$_a$}$$

$$\underset{s,u}{C}\,\underset{u,1}{\hat{x}} - \underset{s,1}{f_x} = 0, \tag{1)$_b$}$$

式中未知量($\underset{n,1}{V}$ 和 $\underset{u,1}{\hat{x}}$)的个数$(n+u)$总是大于方程式的个数$(c+s)$，解不惟一. 根据最小二乘原理，应在"$V^{\mathrm{T}}PV=\min$"准则下求得一组惟一解，这是一个求条件极值的问题. 按条件极值法，需组成如下函数：

$$\Phi = V^{\mathrm{T}}PV - 2K^{\mathrm{T}}(AV + B\hat{x} - f) - 2K_s^{\mathrm{T}}(C\hat{x} - f_x), \tag{2}$$

式中 K 和 K_s 分别是对应于$(1)_a$和$(1)_b$的拉格朗日乘数向量，在测绘学中则称它们为联系数. 函数 Φ 分别对变量 V 和 $\hat{x}$ 求一阶导数，并令其为零，得到两个方程，连同条件方程$(1)_a$和$(1)_b$，即构成总法方程(参见"法方程")，由此即可求得满足条件方程的一组惟一解.

闭合差(closure)　测量平差中常用的一个术语. 它是指某个量的观测结果与其应有值之间的差值. 在某几个量构成几何或物理条件方程的情况下，由于这些量的观测值中包含有误差，它们不能满足方程而产生一定的差值，称此差值为条件闭合差，简称闭合差.

独立观测平差(adjustment of uncorrelated observations)　测量平差的类型之一. 是指对(随机)不相关观测值进行平差的方法. 在平差计算中，如果参与平差的观测值向量，其协方差阵(包括协因数阵和权阵)为对角阵，即观测值之间彼此不相关时，称为独立观测平差(参见"相关观测平差").

相关(观测)平差(adjustment of correlated observations)　简称相关平差. 测量平差的类型之一. 是指对(随机)相关观测值进行平差的方法. 在平差计算中，如果参与平差的观测值向量的协方差阵(包括协因数阵和权阵)为非对角方阵，即观测值之间彼此相关，它们的协方差全部或部分不等于零时，称为相关(观测)平差.

自19世纪初创立最小二乘法以来，在测量平差计算中只局限于根据(随机)独立的观测值确定未知参数的最佳估值. 1947年，钦斯特拉(Tienstra，J. M.)提出了相关平差理论，不仅使得独立的直接观测值可以作为平差的对象，而且它的导出量，例如独立观测值的函数或任何一种初步平差的结果都可作为平差的对象，从而大大扩展了测量平差的应用范围，推动了测量平差理论的发展.

条件平差(condition adjustment)　平差计算中的基本方法之一. 它是根据观测量之间所构成的条件方程式，按最小二乘原理，求得各观测值的估值并进行精度评定的一种平差方法. 条件平差函数模型的线性形式为

$$\underset{r,n}{A}\,\underset{n,1}{V} - \underset{r,1}{f} = 0,$$

式中 $\underset{r,n}{A}$ 是条件方程的已知系数阵，$\underset{n,1}{V}$ 是观测值的改正数向量，$\underset{r,1}{f}$ 是条件方程的常数项，r 和 n 分别表示多余观测数和观测值总数. 在 $V^{\mathrm{T}}PV=\min$ 的条件下解出 V 向量，从而求得观测向量的估值

$$\hat{L}=L+V.$$

在有些平差问题中，由于条件方程式的类型多样、形式不一、无明显规律性，因此建立条件方程式的工作较为复杂，这是条件平差法的不足之处.

附有未知数的条件平差(condition adjustment with unknowns)　平差计算中的基本方法之一．它类似于条件平差法．不同之处是：平差时除了观测量之外，还选定 u 个独立的非观测量作为未知数参与平差计算．所谓独立，是指这些未知数之间不存在任何确定的函数关系．附有未知数的条件平差的线性函数模型为

$$\underset{c,n}{A}\underset{n,1}{V}+\underset{c,u}{B}\underset{u,1}{\hat{x}}-\underset{c,1}{f}=0,$$

上式称为附有未知数的条件方程，式中 $\underset{c,n}{A}$ 和 $\underset{c,u}{B}$ 都是已知的系数阵，$\underset{n,1}{V}$ 和 $\underset{u,1}{\hat{x}}$ 分别是观测值和未知数近似值 X^0 的改正数向量，$\underset{c,1}{f}$ 是方程的常数项．方程式的个数 $c=r+u$，r 是多余观测数，u 是独立未知数的个数．在 $V^TPV=\min$ 的条件下，解出 V 和 $\hat{x}$，从而求得估值 $\hat{L}=L+V$ 和 $\hat{X}=X^0+\hat{x}$.

附有未知数的条件平差法常应用于下述情况：

1. 除了观测量之外，还需要通过平差计算求得某些非观测量的估值和精度．

2. 若不加入未知数，某些条件方程式就难以列出．

3. 为了达到其他目的，例如加入未知数，使整个平差计算达到分区计算的目的等．

参数平差(parameter adjustment)　亦称间接平差．平差计算中的基本方法之一．在进行平差时，选定 t 个独立的非观测量作为参数，并将 n 个观测量逐个表达成这些参数的函数，称为观测方程，t 是表示必要观测值的个数．所谓独立，是指在这 t 个参数之间不存在任何确定的函数关系．经线性化后，其函数模型的形式为

$$\underset{n,1}{V}=\underset{n,t}{B}\underset{t,1}{\hat{x}}-\underset{n,1}{f},$$

上式称为误差方程．式中 $\underset{n,t}{B}$ 是已知的系数阵，它的秩为 $R(B)=t$，即 B 为列满秩；$\underset{n,1}{V}$ 和 $\underset{u,1}{\hat{x}}$ 分别是观测值 L 和参数近似值 X^0 的改正数向量；$\underset{n,1}{f}$ 是误差方程的常数项．在 $V^TPV=\min$ 的条件下，解出 $\hat{x}$ 和 V，从而求得估值 $\hat{X}=X^0+\hat{x}$ 和 $\hat{L}=L+V$.

在参数平差中，往往使所选的参数就是所需要的最后计算成果，例如在水准网中选水准点的高程、在三角网中选待定点的坐标等作为参数．平差工作结束后，就可得到它们的最后结果(估值)和精度．此外，由于误差方程的形式常具有较明显的规律性、编制电算程序较为简便等优点，因此参数平差是当前最常用的一种平差方法．

间接平差(indirect adjustment)　即"参数平差"．

附有限制条件的参数平差(parameter adjustment with constraints)　亦称附有条件的间接平差．平差计算中的基本方法之一．在进行平差时，选定 u 个非观测量作为参数，其中必须包含 t 个(等于必要观测值的个数)独立的参数，其余 s 个是不独立的参数，即在 u 个参数之间存在着 s 个确定的函数关系．这里 $u=t+s$.

该法类似于参数平差，不同之处是：除了将 n 个观测量表达成参数的函数外，还需列出参数之间存在的 s 个函数关系式，称为限制条件方程(或称约束条件方程)．线性化后的函数模型为

$$\underset{n,1}{V}=\underset{n,u}{B}\underset{u,1}{\hat{x}}-\underset{n,1}{f},\tag{1}$$

$$\underset{s,u}{C}\underset{u,1}{\hat{x}}-\underset{s,1}{f_x}=0.\tag{2}$$

(1)式是误差方程，(2)式是限制条件方程．$\underset{n,u}{B}$ 和 $\underset{s,u}{C}$ 是已知的系数阵，$\underset{n,1}{V}$ 和 $\underset{u,1}{\hat{x}}$ 分别是观测值和未知参数近似值 X^0 的改正数向量，$\underset{n,1}{f}$ 和 $\underset{s,1}{f_x}$ 分别是误差方程和限制条件方程的常数项．在 $V^TPV=\min$ 的条件下，解出 V 和 $\hat{x}$，从而求得估值

$$\hat{L}=L+V\ \text{和}\ \hat{X}=X^0+\hat{x}.$$

附有条件的间接平差(indirect adjustment with condition)　即"附有限制条件的参数平差"．

附有限制条件的条件平差(condition adjustment with constraints)　平差计算中的方法之一．在进行平差时，除了 n 个观测量之外，还选定 u 个非观测量作为参数，其中包含有 s 个不独立的参数，即在这 u 个参数之间存在着 s 个确定的函数关系．附有限制条件的条件平差的线性函数模型为

$$\underset{c,n}{A}\underset{n,1}{V}+\underset{c,u}{B}\underset{u,1}{\hat{x}}-\underset{c,1}{f}=0,\tag{1}$$

$$\underset{s,u}{C}\underset{u,1}{\hat{x}}-\underset{s,1}{f_x}=0.\tag{2}$$

(1)式是附有参数的条件方程，(2)式是反映参数之间函数关系的限制条件方程．A,B,C 是已知的系数阵，$\underset{n,1}{V}$ 和 $\underset{u,1}{\hat{x}}$ 分别是观测值和参数近似值 X^0 的改正数向量，$\underset{c,1}{f}$ 和 $\underset{s,1}{f_x}$ 是方程的常数项．

在平差计算时，在选定非观测量作为参数的各种平差方法中，对于所选的参数常有一定的限制和要求．例如，在附有未知数的条件平差中，要求 u 个参数必须是(函数)独立的，且需 $u<t$ (t 表示必要观测值的个数)；在参数平差中，要求 u 个参数必须是独立的，且需 $u=t$；在附有限制条件的参数平差中，要求在 u 个参数中必须包含 t 个独立的参数，且需 $u>t$．但在附有限制条件的条件平差中，对于参数的个数 u，则无任何限制，因此它可以解决不符合上述各种限制和要求的一些平差问题．此外，由于各种平差方法的函数模型都是(1)，(2)式中函数模型的特例(参见"函数模型")，故称该模型为概括平差模型或通用平差模型，由它所导出的平差公式则称为通用公式．

法方程(normal equation)　测量平差计算中的

一个重要的方程.它是由平差函数模型线性化后的方程式,以及在 $V^{T}PV=\min$ 的条件下所导得的方程式合并而成的联立方程组.在各种平差函数模型中,方程式的个数总是少于所求未知量的个数,是多解方程,解不惟一,通过法方程则可求得既满足所有方程式,又满足 $V^{T}PV=\min$ 的一组惟一解.例如,就附有限制条件的条件平差法而言,方程组

$$AV + B\hat{x} - f = 0, \quad (1)$$

$$C\hat{x} - f_x = 0, \quad (2)$$

$$PV - A^{T}K = 0, \quad (3)$$

$$B^{T}V + C^{T}K_s = 0 \quad (4)$$

就是它的总法方程,或称基础方程.其中(1),(2)两式是其函数模型线性化后的方程式,(3),(4)两式是在 $V^{T}PV=\min$"的条件下按条件极值导出的方程式.K 和 K_s 是求条件极值时分别对应于(1)式和(2)式的拉格朗日乘数向量,在测量中习惯称为联系数向量.若利用(3)式消去 V,则得

$$\begin{aligned} N_{aa}K + B\hat{x} \qquad\qquad - f &= 0, \\ B^{T}K \qquad + C^{T}K_s \qquad &= 0, \\ C\hat{x} \qquad - f_x &= 0, \end{aligned} \quad (5)$$

(5)式称为法方程,其中 $N_{aa}=AQA^{T}$.其他各种平差方法的总法方程和法方程都可由(1)~(4)式和(5)式简化而得.

各种平差方法的法方程,其系数阵都是对称满秩方阵.但在选取与定位有关的参数进行参数平差时,若网中无起算数据或起算数据不足,误差方程的系数阵列不满秩,即产生基准秩亏,则法方程的系数阵将是奇异的(参见"秩亏网平差").

总法方程(total normal equation)　见"法方程".

分组平差(adjustment in groups)　平差计算中的方法之一.它是用于处理大型平差问题的一种计算方法.所谓分组,是指将一个平差问题中的全部方程式分成若干个方程组,然后分组进行计算.它的基本要求是:分开计算的结果应与不分开的整体平差的结果完全相等.以附有未知数的条件平差为例,设将全部方程分成 k 组:

$$A_1V + B_1\hat{x} - f_1 = 0, \quad (1)$$

$$A_2V + B_2\hat{x} - f_2 = 0, \quad (2)$$

$$\cdots\cdots$$

$$A_kV + B_k\hat{x} - f_k = 0. \quad (k)$$

平差计算的基本思路是:

1.先对第1组单独进行平差,求出第一次平差值 $\hat{\bar{L}}'$ 及其权逆阵 $Q_{\hat{\bar{L}}'}$:

$$\hat{\bar{L}}' = \begin{pmatrix} \hat{L}' \\ \hat{X}' \end{pmatrix}, \quad Q_{\hat{\bar{L}}'} = \begin{pmatrix} Q_{\hat{L}'} & Q_{\hat{L}'\hat{X}'} \\ Q_{\hat{X}'\hat{L}'} & Q_{\hat{X}'} \end{pmatrix}.$$

2.以前一次的平差值 $\hat{\bar{L}}'$ 作为新的观测值,计算第2组方程的常数项,并以 $Q_{\hat{\bar{L}}'}$ 作为其权逆阵,再对第2组方程单独进行平差,以求出第二次平差值 $\hat{\bar{L}}''$ 和它的权逆阵 $Q_{\hat{\bar{L}}''}$.

3.按同样方法逐组进行平差,直到第 k 组平差结束,以求出最后的平差值 $\hat{\bar{L}}^k$ 和它的权逆阵 $Q_{\hat{\bar{L}}^k}$:

$$\hat{\bar{L}}^k = \begin{pmatrix} \hat{L}^k \\ \hat{X}^k \end{pmatrix}, \quad Q_{\hat{\bar{L}}^k} = \begin{pmatrix} Q_{\hat{L}^k} & Q_{\hat{L}^k\hat{X}^k} \\ Q_{\hat{X}^k\hat{L}^k} & Q_{\hat{X}^k} \end{pmatrix}.$$

上述结果将与不分组的整体平差的结果相同.

分组平差对于下述情况将是一种有效的措施:

1.处理大型平差问题时,因分组平差的法方程的阶数较低,可以克服不分组时计算机容量不敷应用的困难.

2.在一定情况下,例如分区平差,各组可同步进行计算,从而可以缩短计算的周期(参见"分区平差").

3.当一个测量网已经完成了平差计算之后,因网形扩展或加密而增加了新的观测数据时,只需将原平差结果作为相关观测值并顾及其权逆阵,与新观测值一并进行平差,即可求得与新旧资料整体平差的相同结果.特别是在具有规律性强的递推公式的情况下,不仅计算简便,且可及时提供平差的结果,如序贯平差.

分区平差(block adjustment)　平差计算中的方法之一.它是利用法方程系数阵的稀疏性结构而形成的一种分组平差方法.所谓分区是指将一个大型测量网划分成若干个独立的部分,每个部分即称为一个区.一般而言,如果整体平差时的法方程具有如下矩阵结构形式:

$$\begin{aligned} N_aX_a \qquad\qquad + N_{a\lambda}Y + W_a &= 0, \\ N_bX_b \qquad + N_{b\lambda}Y + W_b &= 0, \\ \cdots\cdots \\ N_kX_k + N_{k\lambda}Y + W_k &= 0, \\ N_{a\lambda}^{T}X_a + N_{b\lambda}^{T}X_b + \cdots + N_{k\lambda}^{T}X_k + N_{\lambda}Y + W_{\lambda} &= 0, \end{aligned}$$

即可采用分区平差的方法进行计算.上式中前 k 个矩阵方程分别与各个分区相对应,最后一个方程称为联系方程.$X_a,X_b,\cdots,X_k$ 称为内部(或区域性)未知数,Y 称为公共未知数.

分区平差的实际计算步骤是:

1.分区后分别列出其方程式并组成法方程,消去内部未知数,得到各分区的约化法方程.

2.将各区的约化方程的系数和常数项对应相加,得到只包含公共未知数 Y 的联系约化方程,并由它解出 Y.

3.将得出的 Y 值分别回代到各分区的约化方程,求出其内部未知数 $X_a,X_b,\cdots$.

从理论上可以证明,这样计算的结果将与整体平差的结果完全相等.

分区平差的优点在于：除了联系约化方程的组成和计算之外，其余计算工作均可由各分区同步进行．这样既可缩短整体平差的计算周期，又可解决整体平差时计算机容量不足的困难．但在分区平差中，如果增加了新的观测值或未知数，则需重新进行联系未知数的计算，并改正所有各分区内部未知数的解，这是分区平差的一个主要缺点．

序贯平差(successive adjustment)　平差计算中的方法之一．它是具有规律性强的递推计算公式的一种分组平差方法．以参数平差为例，设有观测向量 L_{k-1} 和 L_k，由它们列出的误差方程相应为

$$V_{k-1}=B_{k-1}\hat{x}-f_{k-1},\quad Q_{k-1}$$
$$V_k=B_k\hat{x}-f_k,\quad Q_k$$

Q_{k-1} 和 Q_k 分别是 L_{k-1} 和 L_k 的权逆阵，下标 k 表示分组的序号($k=2,3,\cdots$)．当完成了第($k-1$)组误差方程的平差计算之后，可以求出第($k-1$)次的参数估值 $\hat{x}_{k-1}$ 和它的权逆阵 $Q_{\hat{x}_{k-1}}$．然后，应用下述递推公式：

$$\hat{x}_k=\hat{x}_{k-1}+Q_{\hat{x}_{k-1}}B_k^{\mathrm{T}}(Q_x+B_kQ_{\hat{x}_{k-1}}B_k^{\mathrm{T}})^{-1}(f_k-B_k\hat{x}_{k-1}),$$
$$Q_{\hat{x}_k}=Q_{\hat{x}_{k-1}}-Q_{\hat{x}_{k-1}}B_k^{\mathrm{T}}(Q_k+B_kQ_{\hat{x}_{k-1}}B_k^{\mathrm{T}})^{-1}Q_{\hat{x}_{k-1}},$$

即可算出第 k 次平差后的参数估值 $\hat{x}_k$ 和它的权逆阵 $Q_{\hat{x}_k}$．

序贯平差的优点是：

1．在进行第 k 组平差时，除了需要保留前一次平差求得的 $\hat{x}_{k-1}$ 和 $Q_{\hat{x}_{k-1}}$ 外，不需要保留以前的观测数据，从而大大节省了计算机的存储空间．

2．在递推公式中需计算逆矩阵

$$(Q_k+B_kQ_{\hat{x}_{k-1}}B_k^{\mathrm{T}})^{-1},$$

该矩阵的阶数只取决于第 k 组观测数据的个数．如果新增加的观测值个数不多，则只需计算一个低阶矩阵的逆．特别是当只增加一个观测值时，则变成计算一个纯量的倒数．因此，整体平差中高阶法方程的求逆就被一系列简单的矩阵运算所代替．

3．序贯平差是以一种反复的、动态的方式逐次进行运算的，特别适用于参数实时的连续估计，以便能及时提供平差的结果和对数据采集进行实时的质量控制．

秩亏网平差(rank defect network adjustment)　亦称自由网平差．平差计算中的方法之一．它是当测量网选取与定位有关的参数进行参数平差时，其误差方程式的系数阵为列不满秩的一种平差方法．引起系数阵列不满秩的原因有：

1．网中的必要观测数据不足，使网的形状无法确定．

2．网中无必要的起算数据，使网的位置、方位和尺度无法确定．

实际工作中，在测量网中总是要进行多余的观测，因此通常所称的"秩亏网"都是指无必要起算数据的测量网．以测量网点坐标为参数的误差方程为

$$\underset{n,1}{V}=\underset{n,u}{B}\underset{u,1}{\hat{x}}-\underset{n,1}{f}.$$

由于网中无起算数据，$R(B)=t<u$，即 B 的秩等于必要观测值的个数 t，而小于未知参数的个数 u，其秩亏数 d $(=u-t)$ 等于必要起算数据的个数．例如，在水准网中 $d=1$，在平面测边网、边角网中 $d=3$，在平面测角网中 $d=4$ 等．

在 $V^{\mathrm{T}}PV=\min$ 的条件下可导出求解 $\hat{x}$ 的法方程，由于 B 列不满秩，法方程的系数阵将是 u 阶的奇异方阵，其解不惟一．为了获得一组惟一解，还必须加入新的约束条件，即解向量 $\hat{x}$ 的最小范数条件

$$\hat{x}^{\mathrm{T}}\hat{x}=\min \text{ 或 } \hat{x}^{\mathrm{T}}P_x\hat{x}=\min.$$

按这些条件求解参数估值的方法分别称为普通秩亏网平差和加权秩亏网平差．有时还可在部分解向量的范数为最小的条件下计算，如拟稳平差．

自由网平差(free network adjustment)　即"秩亏网平差"．

拟稳平差(quasi-stable adjustment)　秩亏网平差的方法之一．在无足够起算数据的测量网中，当选取与定位参数有关的参数进行参数平差时，其误差方程的系数阵列不满秩．拟稳平差的做法是：将网中所有待定点分成两部分，一部分是相对于其他点而言稳定程度较好的点，称为稳定点；另一部分称为非稳定点．设非稳定点和稳定点的坐标改正数向量分别为 x_{I} 和 x_{II}，在进行平差时，为了求得一组惟一解，除了遵循最小二乘准则"$V^{\mathrm{T}}PV=\min$"外，还需加入稳定点坐标改正数的范数最小条件 $\hat{x}_{\mathrm{I}}^{\mathrm{T}}\hat{x}_{\mathrm{I}}=\min$，因此，它是在"部分解向量范数最小"条件下的一种求解方法．周江文对拟稳平差进行了深入的研究，在这一研究领域中起到了先导和推动的作用．

误差椭圆(error ellipse)　用图形描述点位误差的一种方法．即描述待定点在各个方向上误差分布的规律．在通常情况下，平差后一般只求出测量网中待定点在 x 和 y 两个坐标轴方向上的中误差(亦称位差)．但在有些测量问题中，除了要估计待定点在两个坐标轴方向上的位差之外，还要了解在其他方向上的误差分布情况，为此，要做出该待定点上的误差椭圆．

确定某待定点上误差椭圆的方法是：设已知该点坐标平差值的协因数和互协因数分别为 Q_X，Q_Y 和 Q_{XY}，则可由下式求得其误差椭圆的长半轴 E 和短半轴 F：

$$\begin{cases}E^2=\dfrac{1}{2}\hat{\sigma}_0^2\{(Q_X+Q_Y)+K\},\\ F^2=\dfrac{1}{2}\hat{\sigma}_0^2\{(Q_X+Q_Y)-K\},\end{cases}\tag{1}$$

式中 $\hat{\sigma}_0^2$ 为单位权方差的估值，而

$$K = \sqrt{(Q_{\hat{X}} - Q_{\hat{Y}})^2 + 4Q_{\hat{X}\hat{Y}}^2} \tag{2}$$

(K 永远取正值). 同时由

$$\tan 2\varphi_0 = \frac{2Q_{\hat{X}\hat{Y}}}{Q_{\hat{X}} - Q_{\hat{Y}}} \tag{3}$$

可以解出两个极值方向为 φ_0 和 $\varphi_0+90°$,两者相互正交. 若 $Q_{\hat{X}\hat{Y}}$为正,则长半轴 E 应在第一、三象限,短半轴 F 应在第二、四象限;若 $Q_{\hat{X}\hat{Y}}$为负,则反之. 例如,已知三角网中某待定点 P 的坐标协因数为 $Q_{\hat{X}}=2.10$, $Q_{\hat{Y}}=1.60$, $Q_{\hat{X}\hat{Y}}=-0.25$(单位均为 $\mathrm{dm}^2/('')^2$),单位权方差 $\hat{\sigma}^2=1.0\,('')^2$,由(3)式得 $\tan 2\varphi_0=-1$,故 $2\varphi_0=135°$或 $315°$,即 $\varphi_0=67.5°$或 $157.5°$. 因 $Q_{\hat{X}\hat{Y}}$为负,故知长半轴 E 应在二、四象限,即 $\varphi_E=157.5°$或 $337.5°$,$\varphi_F=67.5°$或 $247.5°$. 由(1)和(2)式得

$$K=0.71,\ E^2=2.20\mathrm{dm}^2,\ F^2=1.50\mathrm{dm}^2,$$

即

$$E=\pm 1.48\mathrm{dm},\quad F=\pm 1.2\mathrm{dm}.$$

根据这些数据即可做出 P 点上的误差椭圆,如图所示.

对于空间点,则有误差椭球. 此时可将上述方法从二维推广到三维.

权函数(weight function) 为精度评定而建立的一种函数. 在平差计算中,有时为了评定某个量或多个量平差后的精度,就要将这些量表达成观测值平差值和参数平差值的函数,即

$$\hat{\varphi} = \hat{\Phi}(\underset{n,1}{\hat{L}}, \underset{u,1}{\hat{X}}),$$

式中 $\hat{\varphi}$ 表示所求量的平差值. 由于它是为评定 $\hat{\varphi}$ 的精度而建立的函数,故称其为权函数. 当权函数为非线性形式时,还需求其全微分,得

$$\begin{aligned} \mathrm{d}\hat{\varphi} &= k_1\mathrm{d}\hat{L}_1 + k_2\mathrm{d}\hat{L}_2 + \cdots + k_n\mathrm{d}\hat{L}_n \\ &\quad + k_{x_1}\mathrm{d}\hat{X}_1 + k_{x_2}\mathrm{d}\hat{X}_2 + \cdots + k_{x_u}\mathrm{d}\hat{X}_u \\ &= K\mathrm{d}\hat{L} + K_x\mathrm{d}\hat{X}, \end{aligned}$$

其中

$$k_i = \frac{\partial\Phi}{\partial\hat{L}_i}\ (i=1,2,\cdots,n),$$

$$k_{x_j} = \frac{\partial\Phi}{\partial\hat{X}_j}\ (j=1,2,\cdots,u).$$

然后利用平差后的协因数阵,如 $Q_{\hat{L}}$, $Q_{\hat{X}}$, $Q_{\hat{L}\hat{X}}=Q_{\hat{X}\hat{L}}^{\mathrm{T}}$等,应用协因数传播律即可求得 $\hat{\varphi}$ 的权倒数(即协因数)$Q_{\hat{\varphi}}$,并进而求得其方差 $\sigma_{\hat{\varphi}}^2=\hat{\sigma}_0^2Q_{\hat{\varphi}}$,$\hat{\sigma}_0^2$ 为单位权方差的估值.

权系数阵(weight coefficient matrix) 测量平差精度评定中的一个重要概念. 即在参数平差中,法方程系数阵的逆阵即为参数平差值的权逆阵. 由最小二乘平差知,参数平差中的误差方程和法方程分别为

$$\underset{n,1}{V} = \underset{n,t}{B}\underset{t,1}{\hat{x}} - \underset{n,1}{f} \quad 和 \quad \underset{t,t}{N_{bb}}\underset{t,1}{\hat{x}} - \underset{t,1}{B^{\mathrm{T}}Pf} = 0.$$

且知,平差后参数平差值的协因数阵为

$$Q_{\hat{X}} = N_{bb}^{-1},$$

式中 $N_{bb}=B^{\mathrm{T}}PB$,即法方程系数阵的逆阵就是参数平差值的协因数阵,该矩阵为 t 阶对称方阵,其中 t 个主对角元素分别是 t 个参数的权倒数,非主对角元素则是相应两个参数的相关权倒数. 由此可知,$Q_{\hat{X}}$ 中的元素都直接是参数的权倒数或相关权倒数,权系数阵即因此而得名.

协方差函数(covariance function) 测量平差中的一个重要概念. 它是用以确定协方差阵中各元素大小的一种函数. 例如,在最小二乘滤波、推估、配置中,具有随机性质的信号,其协方差阵中的元素就是应用事先求得的协方差函数来计算的.

协方差函数一般是某些变量的函数. 要确定它,需要具有大量的试验数据,因此要确定一个协方差函数的具体形式,是一个相当复杂而困难的问题. 但在实际工作中,往往是提出某些假设,使得确定这种函数的工作既简单又基本符合客观实际. 例如,通常假定信号是具有各态历经性的平稳随机函数,在这种假设下,就可以通过一次试验的样本来确定其协方差函数,而且它变成一个变量的函数. 例如,由希尔伏宁(Hirvonen,R. A.)根据美国俄亥俄州空间异常所确定的协方差函数为

$$\sigma(d) = \frac{\sigma(0)}{1+\left(\dfrac{d}{d_0}\right)^2},$$

式中 $d_0=40\mathrm{km}$,$\sigma(0)=337\mathrm{mgal}^2$(参见"重力"),$d$ 为两点间的公里数,上式适用于 $d<100\mathrm{km}$ 的范围.

最小二乘滤波(least squares filtering) 亦称无系统参数的最小二乘配置. 由观测数据确定随机参数最佳估值的一种方法. 即在其函数模型中不包含非随机参数的部分(参见"最小二乘配置").

滤波原是无线电工程中的一个术语. 它本来的含义是从接收到的受干扰的电磁波信号中,排除噪声的干扰,分离出所需要的信号. 在一般的数据处理问题中,滤波就是通过含有误差(噪声)的观测数据,确定随机参数(信号)的最佳估值的方法. 最小二乘滤波的线性函数模型为

$$-V + \bar{A}V_{\bar{S}} - f = 0,$$

式中 $-V$ 是观测误差(即噪声)的估值,

$$V_{\bar{S}} = \begin{bmatrix} V_S \\ V_{S'} \end{bmatrix},$$

V_S 是观测点信号 S 的改正数向量,$V_{S'}$ 是未测点信

号 S' 的改正数向量，f 是常数项，$\bar{A}=[A\ 0]$ 是已知的系数阵.

信号 S 和 S' 是随某种因素变化的随机函数向量，用以描述它们随机性质的协方差阵，其元素的大小可由协方差函数确定. 根据广义最小二乘准则

$$V^{T}P_{n}V+V_{\bar{S}}^{T}P_{\bar{S}}V_{\bar{S}}=\min$$

解出 V 和

$$V_{\bar{S}}=\begin{bmatrix}V_{S}\\V_{S'}\end{bmatrix},$$

从而求得信号的估值 $\hat{S}=S+V_{S}$ 和 $\hat{S}'=S'+V_{S'}$. P_{n} 和 $P_{\bar{S}}$ 分别是观测误差(噪声)和信号的先验权阵.

最小二乘配置(least squares collocation) 亦称最小二乘拟合推估. 由观测数据同时确定随机参数和非随机参数估值的一种方法. 它是包含最小二乘平差、滤波和推估的一种更广义的平差方法. 平差是指确定非随机参数的估值，滤波是指确定观测点的信号估值，推估是指确定未测点上的信号估值. 最小二乘配置的线性函数模型为

$$-V+\bar{A}V_{\bar{S}}+B\hat{x}-f=0,$$

式中 $-V$ 是观测误差(又称噪声)的估值，

$$V_{\bar{S}}=\begin{bmatrix}V_{S}\\V_{S'}\end{bmatrix},$$

V_{S} 是观测点信号 S 的改正数向量，$V_{S'}$ 是未测点信号 S' 的改正数向量，f 是常数项，$\bar{A}=[A\ 0]$ 和 B 是已知的系数阵.

信号和噪声是通信工程学中的术语，它们都是随机量. 不同于一般的随机变量，信号 S 和 S' 是在实验过程中随着某种因素(例如时间、距离等)的变化而连续变化的随机量，称为随机函数. 用以描述信号随机性质的协方差阵，其元素的大小是由与某些因素有关的函数确定，这种函数称为协方差函数. 根据广义最小二乘准则

$$V^{T}P_{n}V+V_{\bar{S}}^{T}P_{\bar{S}}V_{\bar{S}}=\min$$

解出 V 和

$$V_{\bar{S}}=\begin{bmatrix}V_{S}\\V_{S'}\end{bmatrix},$$

从而求得信号的估值 $\hat{S}=S+V_{S}$ 和 $\hat{S}'=S'+V_{S'}$. P_{n} 和 $P_{\bar{S}}$ 分别是观测误差(噪声)和信号的先验权阵.

最小二乘拟合推估(least squares collocation) 即"最小二乘配置".

测量器具

指南针(magnetic needle) 测定南北方向的一种器具. 它是中国古代重大发明之一，是一种最古老的指示方向的测量仪器. 它的主要部分是一根能绕轴转动的磁针，磁针在地磁场的作用下始终保持在磁子午线方向上，可据此迅速判别南北方向. 目前，指南针除在旅行、行军等方面应用外，还按其原理制成了各种仪器，如地质工作中用的地质罗盘、测绘工作中用的罗盘仪、航空和航海中用的磁罗经等.

司南(southward pointer) 汉墓出土文物之一. 是中国古代创造的指示方向的仪器. 它由方形底盘和磁匙两部分组成. 底盘用铜制成，上面刻有方位；磁匙司南用天然磁石制成，匙底做成球形，置于底盘中央. 由于两者接触面很小，且磁匙两端保持平衡，故磁匙在底盘上可以自由转动，使匙柄始终指向磁北.

指南车(southward pointing cart) 亦称司南车. 是中国古代利用机械原理制造的一种指示方向的装置，其外表像马车，车上站立一木人，车的方向随意改变，但木人的手始终指向南方.

司南车(southward pointing cart) 即"指南车".

记里鼓车(li-drum-hodometer) 中国古代用于记录距离的一种装置. 外形像马车，车上装一大鼓，鼓的两边各站一持锤木人. 车前进时，车轮转动的距离通过齿轮系统传给木人. 车轮行一里(1 里=0.5km)，木人敲鼓一次.

罗盘仪(compass) 测定磁方向角的仪器. 主要由磁针、度盘、照准器等部件构成. 度盘刻度有全周式和象限式两种方式. 使用时将罗盘仪水平置于测站上，用照准器瞄准所测方向，此时磁针在度盘上的读数，即为该方向线的磁方位角或磁象限角. 若测站点上观测两个方向，则该两方向的磁方位角之差，即为此二方向线间之水平角. 测量作业中常用有支架的罗盘仪；地质工作者则常用带有测量岩石倾角等附件的罗盘仪.

挂罗盘(hanging compass) 17 世纪中叶，罗斯勒(Rossler)发明了适用于矿山井下测量的仪器. 目前普遍使用的是其改良型，主要有阿斯特兰地挂罗盘、李倚次尔德挂罗盘、潘寇德挂罗盘等. 挂罗盘测量精度会受局部引力的影响(由于磁针附近存在铁器、电流或磁性矿物等，以致磁针不指向磁北). 当检查出有局部引力影响时，应按有关规定和方法加以更正.

经纬仪(theodolite) 一种应用最广泛的测量仪器，主要用于测量水平角和垂直角. 由望远镜、水平度盘、垂直度盘、水准器和基座等构成. 望远镜内一般都装有十字丝用于照准方向，多数经纬仪还附有视距装置，可用来测定距离. 它的种类很多，构造则大同小异. 按精度分，有普通经纬仪和精密经纬仪；按用途分，有天文经纬仪、大地经纬仪、矿山经纬仪、摄影经纬仪等；还有些特殊性能的经纬仪，如自动穿孔记录度盘读数的编码经纬仪、自动连续照准

空中目标的跟踪经纬仪、自动保持目标方向的陀螺经纬仪和激光经纬仪等；按读数设备分，还有游标经纬仪和光学经纬仪．国产经纬仪系列标准有 DJ07，DJ1，DJ2，DJ6，DJ15 和 DJ60 六种型号．其中 D，J 分别为“大地测量”、“经纬仪”的汉语拼音第一个字母，07，1，2，6，15，60 分别表示该仪器以秒表示的一个测回水平方向的中误差．

摄影经纬仪(phototheodolite) 地面摄影测量野外作业用的仪器．主要由摄影机和经纬仪两部分组成．两者有固定在一起的，也有分开的，后者在摄影机上另备有定向装置．这类摄影机的内方位元素是已知的，不能调焦，主要用于地形测量．供地形测量和近景摄影测量两用的摄影机，称为地面全能摄影机．

平板仪(plane-table) 地形测量用的一种仪器．主要由照准仪、测图板、三角架、基座、水准器、对点器和方框罗针等组成．照准仪带望远镜和竖直度盘的称为大平板仪，也称为平台仪．用望远镜照准竖立在测点上的标尺，根据视距和竖角读数可计算出测站至尺点的水平距离和高差．仅由一固定长度木尺的两端安置两块觇板组成的照准仪，称为小平板仪，亦称照准测斜仪．后者轻便、快速，但精度低，在碎部测量中，常将它与经纬仪配合使用．平台仪分一般平台仪和自动归算平台仪两类．

照准测斜仪(peepsight alidade) 见“平板仪”．

电子平板(electronic plane-table) 便携式微机及有关软件构成的数字测图系统．用于大比例尺地籍图、地形图的野外一体化测图．与电子全站仪在线连结采集数据，通过编码和图形处理软件，在计算机屏幕上生成图形，可在现场进行编辑和修改．随着测绘科技的进步，电子平板将逐渐取代传统的平板仪测图．

水准仪(level) 用于准确测量高差的仪器．主要由望远镜、长水准管和基座组成．利用长水准管使望远镜视准轴水平，配合水准标尺进行测量．按精度分，有普通水准仪和精密水准仪两类；按结构分，有微倾斜水准仪和利用光学与力学结构代替长水准管的自动安平水准仪；按望远镜结构分，有定镜水准仪和活镜水准仪；此外还有激光水准仪等．国产水准仪系列标准有 DS_{05}，DS_1，DS_3，DS_{10}，DS_{20} 等型号．其中 D 和 S 分别为“大地测量”和“水准仪”的汉语拼音第一个字母，数码 05，1，3，10 及 20 为该仪器以毫米为单位，每公里水准测量高差中的偶然中误差．水准仪在大地测量、地形和工程测量方面广泛使用．

真空气压计(aneroid barometer) 测量气压的仪器．现在所使用的气压计有三种：水银气压计、真空气压计、沸点气压计．真空气压计应用最普遍，直径几厘米至十余厘米不等．它是由感应、传递放大和指示三部分组成．其工作原理是：空气压力作用于真空金属小盒的盒壁，通过杠杆传递到指针而得出气压读数．盒上装有弧形温度计，用以测定仪器温度．此外还备有一支温度计供测定空气温度用．真空气压计的精度较低，只作温度改正：将读数乘以$(1+0.0037t)$，式中 t 为沿测线的平均温度(摄氏度)，温度在零度以下改正值为负，温度在零度以上改正值为正．

水准标尺(levelling staff) 简称水准尺．配合水准仪进行水准测量的主要工具．一般长 3m，分普通水准标尺和精密水准标尺两类．普通水准标尺一般为木质，其两面分别喷制黑白和红白相间的厘米分格，由下向上逐次标注长度；精密水准标尺为木质框架的中央装一条因瓦合金带尺，尺上喷制 1cm 或 0.5cm 间隔的分划线，用弹簧引张在尺面上，也称因瓦水准尺．

水准尺(leveling staff) 即“水准标尺”．

视距尺(stadia rod) 地形测量中与视距仪配合使用的标尺．用优质木料制成，尺上最小分划一般为 1cm，也有 2cm、5cm 和 10cm 的．还有在一根标尺上刻有两种不同的分划，用于测定远近不同的距离．

基线尺(baseline measuring apparatus) 丈量三角网(锁)中基线长度或其他精密边长的工具．早期的基线尺是 3m 木制杆尺．20 世纪以来，都用因瓦基线尺，但在低精度测量中，也可用钢线尺或钢卷尺作为基线尺．

因瓦基线尺(invar baseline wire) 测量工作中丈量基线长度的一种精密工具．用温度系数极小的铁镍合金因瓦制作．分线状尺和带状尺两种，前者一般长 24m，后者一般长 50m．

一级线纹米尺(standard meter) 测量工作中精确测量和检验直线长度的直尺．由膨胀系数很小的合金制成．尺上附有温度计和两个可滑动的放大镜．测量范围为 0～1000mm．最小分划为 0.2mm．一般用来检查精密水准尺的分划和坐标网、图廓点、控制点的展绘精度．

石英钟(quartz clock) 一种精密的计时仪器和频率标准．其核心部件是一个应用石英的压电效应原理制成的稳定晶体振荡源．放在恒温器内的石英晶体产生的振荡频率多为 5.0MHz 和 2.5MHz，再用分频电路分为 100kHz、1kHz 及秒脉冲，使同步马达带动钟面或用数字电路推动数字钟面指示时刻．1929 年，石英钟问世以来，经过不断改进，精度大为提高，到 20 世纪 50 年代中期已在天文台内完全代替天文摆钟．目前高精度石英钟的频率日稳定度已达 10^{-10}，即每天计时误差不超过十万分之一秒．

传感器(sensor)　接收和记录电磁波辐射的仪器.专用于遥感技术的传感器,有被动式和主动式两种.各种类型的摄影机、电视摄相机、电荷耦合器件阵列传感器、自动扫描仪等属被动式传感器,而各种类型的雷达则属主动式传感器.摄影机型传感器只能用于可见光、紫外或近红外波段,直接在感光胶片上成像.电视摄相机用光导管成像,经电子束扫描变成电信号输出或记录在磁带上.CCD 阵列传感器通过 CCD 将接收到的光信号变成电信号,进而变成数字影像.扫描仪、辐射仪和雷达系统都是把接收到的电磁波转变成电信号输出或记录在磁带上,经过处理都可以显示成图象和相片.

原子钟(atomic clock)　一种精密的计时仪器和频率标准.应用原子能级跃迁吸收或发射一定频率的电磁波辐射作为频率振荡源而制成.通常由原子频率标准、晶体振荡器伺服电路和石英钟组成.1955 年,英国首先制成铯原子钟.1967 年起,国际上采用铯原子的跃迁频率规定了原子时的秒长,作为当今时间单位的标准.原子钟具有高准确度和高稳定度,得到了广泛的应用,是制导、空间跟踪、甚长基线干涉测量、相对论效应的验证、地球自转不均匀性等研究必备的仪器.目前,原子钟主要有氢钟、铯钟和铷钟.

电磁波测距仪(electro-magnetic distance measuring instrument)　简称 EDMI.以电磁波为载波的测距仪.按载波不同分为微波测距仪、光电测距仪、激光测距仪和红外测距仪四种;按测程远近分为近程、中程和远程三种;按测算原理分为脉冲式、相位式(固定频率和变频两种)和脉冲相位式.在卫星大地测量、大地控制测量、地形测量和工程测量等各领域中已广泛应用,并朝着轻小高效、高精度、全能化、自动化和数字化方向发展.

光电测距仪(electro-optical distance measuring instrument)　电磁波测距仪的一种.是用可见光为光源的测距仪.它只能在夜间工作,1948 年用于大地测量.测程由几公里到几十公里,其距离测算原理参见“激光测距仪”.

激光测距仪(laser distance measuring instrument)　电磁波测距仪的一种.是在光电测距仪(1948 年用于大地测量)基础上发展起来的一种用激光作为光源的测距仪.由于激光的亮度高、方向性强和单色性好,用它作为测距仪的光源,减少了仪器的体积,减轻了仪器的重量,克服了白天不能工作的缺点,提高了测量精度,扩大了测程范围.

采用不同激光光源制成的激光测距仪种类很多,按测程大小可分三类:

1.短距激光测距仪.测程在 5km 以内,适用于各种工程测量.

2.中距激光测距仪.测程为 5 至几十公里,适用于大地测量和地震预报等.

3.远程激光测距仪.用于测量导弹、人造卫星、月球等空间目标的距离.

激光测距是通过测量激光光束在待测距离(D)上往返传播的时间(t)来算出距离的,其表达公式为

$$D=\frac{1}{2}c\cdot t,$$

式中 c 为激光在大气中传播的速度.

根据传播时间 t 的测定方法不同,测距方法可分为下列两类:

1.脉冲式.测定光脉冲在待测距离上往返传播的时间,此法测距精度不高,大都在米的量级.

2.相位式.通过测量连续调制的光波在待测距离上往返传播所发生的相位变化,间接测量时间 t.这种方法精度较高,在大地和工程测量中已广泛应用.

现代的测距仪普遍采用组件和组合电路,向小型、轻便、自动化以及数字化发展,有的已配用小型电子计算机.

1974 年,中国成功研制了 JCY-2 型精密激光测距仪,1975 年又成功研制 DC-30JG 激光测距仪.

电子全站仪(electronic total station)　能同时用光电方法测量角度(水平角、天顶距)和距离的现代地面测量仪器.它同时具有电子经纬仪和电磁波测距仪的功能.根据仪器的结构,又分组合式和整体式两种类型.仪器中微处理器配置不同,其数据处理能力差别很大,其中电脑型全站仪上可运行用高级语言编写的大型软件包,可实现控制、放样和碎部测量数据采集与处理的一体化.电子全站仪具有与微机相似的通信接口和操作面板以及显示屏幕,可与微机、掌上型电脑或其他数据终端进行数据通信.带有步进马达和 CCD 传感器的电子全站仪称为测量机器人或测地机器人,它可以绕竖直轴作 360°旋转,望远镜可绕其水平轴作上、下 90°的旋转,通过无线电信号和反射信号,测量机器人可自动寻找和精确照准目标,若用程序控制,能像人一样地完成某些特定的测量工作,可实现测量的自动化、智能化以及数据采集处理的一体化.电子全站仪的合作组件是各种棱镜,距离测量范围从数米至十多公里乃至更远;测距精度从亚毫米级至厘米级;角度测量精度为 0.5″至数秒,与精密光学经纬仪的精度相当,甚至更高.电子全站仪是对地面测量仪器的一次革命.20 世纪末已实现将全球定位系统接收机与全站仪组合在一起,在大范围内作自由设站并完成各种常规的地面测量任务.

测地机器人(georobot)　见“电子全站仪”.

测量机器人(measurement robot)　见“电子全

站仪”.

激光雷达(laser radar) 亦称光达.使用激光脉冲或连续激光束的雷达装置.其工作波段为可见光或近红外.由于它具有使用的波长短和激光光束发射角很小的优点,较微波雷达的分辨率要高很多,但受大气吸收影响大,只适用于较短距离(一二十公里以内)的测量.

光达(laser radar) 即“激光雷达”.

航空摄影机(aerial camera) 简称航摄仪.一种专用于在飞机或其他飞行器上向地面进行摄影的照相机.主要由镜箱、光阑和快门、胶片暗盒、座架、动力和控制系统等部分组成,此外还配备有检影器、高差仪、滤色镜等附属设备.它具有较高的光学几何精度,主距固定,焦面上有框标可供测量,并能按设定的时间间隔自动进行连续摄影(启闭快门和卷片)等特点.按特性分为几何精度较高的测量航摄仪和解像能力较高的侦察航摄仪.

现代的高精度航空摄影机上还装有前移补偿装置,能自动补偿飞行器前移引起的像点位移.

航摄仪(aerial camera) 即“航空摄影机”.

无线电测高仪(radio altimeter) 利用无线电波测定飞机高度的仪器.航空摄影时,在航摄仪露光的同时,无线电测高仪的发射机向地面发射无线电信号(一般是脉冲波),经地面反射回来被接收机接收.根据从发射到接收的时间间隔和电波传播速度,由指示器指出航摄仪与地面最近点的距离,并自动记录在专门的胶片上.然后,通过摄影测量方法的处理,便可得出摄影站相对于地面的航高.

立体镜(stereoscope) 对立体像对进行立体观察的简单工具.主要由两个透镜构成.透镜的作用是使左眼只看左相片,右眼只看右相片(称为分相),并使影像放大.通过立体镜观察两张相片的影像重叠部分,便能看到立体模型.立体镜分为反光立体镜和袖珍立体镜等型式.前者利用两组与水平面成45°的反光镜,将观测者的眼基距扩大,可观察大像幅的相片;后者因同名像点间距离必须小于或等于眼基距,因此只能观看小尺寸相片.立体镜的结构简单,携带方便,在航空地质和摄影测量等工作中得到广泛应用.

立体量测仪(stereometer) 亦称地形立体量测仪.航空摄影测量中微分法测图用的一种内业仪器.它由基座、滑床、相片盘、视差改正机械、观测系统及测量系统等构成.使用时,将像对安置在相片盘上,根据分布在标准位置的四个高程控制点和一至两个高程检查点进行相片定向.定向的目的是正确安置各个改正机械,使测量的左右视差较为理想像对的数值.将测量点与基准点的左右视差较代入高差公式,即可计算该点与基准点的准确的高差,并在相片上直接描绘等高线.它适用于丘陵地区测图.

地形立体量测仪(stereometer) 即“立体量测仪”.

立体坐标量测仪(stereocomparator) 航空摄影测量中测量立体像对像点的相片平面坐标的仪器.由基座、相片盘、观测系统和照明系统等构成.使用时,像对安置在左右两相片盘上,定向操作后进行立体观测,即可测量出像点在左相片上的平面坐标和视差,经简单换算又可得出同名点在右相片上的平面坐标,用于计算像对相对定向元素,研究对象的变形观测,应用最多的是为解析空中三角测量室内加密测图控制点提供原始的观测数据.有普通型和精密型两类,普通型的测量精度一般为0.01mm,精密型的测量精度多为1～2μm.精密的立体坐标量测仪一般都可与坐标自动记录装置联用.

立体测图仪(stereoplotter) 航空摄影测量中全能法测图仪器的统称,是严格符合“摄影过程的几何反转”原理的精密光学机械型仪器.主要部件有投影系统、观测系统和测图系统等.种类繁多,多数是空间型的,但也有平面型的.按其投影系统结构方式不同,分为光学投影型、光学机械投影型和机械投影型三类.还有利用电子计算装置进行数字投影的,称为解析测图仪.立体测图仪除用于测图外,有的还可用于空中三角测量或处理地面摄影资料.

精密立体测图仪(precise stereoplotter) 航空摄影测量发展过程中最早的立体测图仪.它是20世纪30年代制成的仪器,属光学投影类.仪器结构庞大复杂,除用于测图外,还可在仪器上做空中三角测量,还可将Y-Z坐标对换,供地面立体摄影测量测图用.第二次世界大战后,它被机械投影类型的仪器所代替.由于这类立体测图仪都是精密的光学机械仪器,故从20世纪60年代后不再用“精密”二字.

纠正仪(rectifier) 航空摄影测量中用于平坦地区相片纠正的仪器,即用光学方式使相片实现透视变换的仪器.主要由底片盘、物镜、承影板、光学条件控制器和操作系统等部分构成.现代的都是竖立式的第二类型(即变换光束)的纠正仪.使用时,将底片放在底片盘内,利用纠正仪的各个动作,使底片上的控制点的投影影像与承影板上按坐标展绘出的相应的点位相重合,即可获得一定比例尺的相当于水平相片的影像.再经过晒像、冲洗等过程,即得到一张纠正好的相片.由于这种类型纠正仪不能改正地形起伏引起的相片误差,所以起伏地区的相片纠正还需利用可作微分纠正的正射投影仪.

投影转绘仪(projector for transferring) 亦称单投影器.航空摄影测量的光学图解类纠正仪.在微分法测图中用于从相片上分带投影转绘地貌和地物.它是由基座、投影器、投影支架和升降系统等部

件组成．使用时，将绘有地物、地貌或其他专业内容的航摄相片制成透光的缩小正片，放在投影器内，通过纠正对点，使相片上控制点的投影与承影板上的相应点位重合，然后进行分带投影转绘，把相片上的地物和等高线描绘在图底上，制成地形图或其他专门地图．

单投影器(single projector)　即“投影转绘仪”．

反光转绘仪(reflecting sketchmaster)　航空摄影测量中最简单的光学图解型纠正器．在观察光路中置入一个半透光、半反射的平面反光镜，使之能同时看到相片影像与图底，通过相片的微小的移动及转动，使相片上控制点的影像与承影板上的相应点位重合，然后按相片上的地物影像勾绘到图板上．用于平坦地区地物转绘或地图修测中的转绘．

多倍投影测图仪(multiplex)　简称多倍仪．航空摄影测量中进行光学投影立体测图的一种仪器．由绘图桌(承影台)、支架、按航摄机镜箱相似地缩小的投影器(包括聚光器)、测绘台等部件组成．支架有1m、2m和3m三种，分别悬挂三个、六个、九个投影器．首先用专门的缩小仪，将航摄底片制成缩小的透明正片，放在投影器内，聚光器照明后就得到与摄影时几何关系相似的投影光线束；相邻两相片的投影光束，经过相对定向，达到同名光线对对相交，构成立体模型；再根据野外控制点进行绝对定向，构成与地面相似并符合测图要求的光学立体模型；然后借补色法进行立体观察，用测绘台测绘地物和地貌，以制成地形图．

多倍仪(multiplex)　即“多倍投影测图仪”．

航空摄影飞机(aerial photography craft)　简称航摄飞机．专用于航空摄影的飞机．各种类型的飞机原则上均可改作航摄飞机，但随摄影任务不同，需选用不同类型的飞机．如用于军事侦察摄影，要求能够飞得高、飞得快的飞机；用于地形摄影测量，要求有较高的稳定性，续航时间要长．一般在飞机的底部开一个或几个窗口，供安置航摄仪和其他设备．此外，为使领航员有较宽的视野，领航员的座舱最好在机头的最前部，特殊的大比例尺的航空摄影，也可用直升机．

解析测图仪(analytical plotter)　是利用电子计算机与立体坐标量测仪相连结，用严格数学方法实时计算像点坐标和模型坐标几何关系、模型坐标与地面坐标以及图面坐标几何关系的摄影测量仪器．它既可输出图解产品，也可输出记录在磁介质上的数字产品．

解析测图仪的硬件部分包括：用作立体观测的精密立体坐标量测仪，实现解析测图仪各种实时和非实时计算的电子计算机，在电子计算机控制下自动绘制各种图件的数控绘图仪，以及连结各个部件和信息通信用的接口和控制部件．作业员对解析测图仪的操作除了一般测图仪上的手轮、脚盘外，还可以通过电子计算机的终端、专用面板和“菜单”键盘等实现．

除了电子计算机提供的系统软件外，解析测图仪具有专门为摄影测量目的制作的软件，包括实时程序和应用程序．实时程序体现数字导杆作用，通过必要的坐标变换建立像方和物方几何关系．以30～100Hz/s的频率执行，来保证立体模型的连续和完整．应用程序包括内定向、相对定向、绝对定向、一步定向、模型存储与恢复、点观察、数字高程模型采样、面积和体积及矢量计算、空中三角测量和机助绘图与数字测图．新一代解析测图仪还可进行在线影像匹配、自动化空中三角测量、近景摄影测量和遥感图象的解析处理．

1957年，美国人海拉瓦(Helava，U.)首先发表了解析测图仪原理的论文，被认为是解析测图仪的发明者．经过近20年的研究试验阶段，1976年，在赫尔辛基召开的第十三届国际摄影测量大会上展出了七家公司的解析测图仪，并转入市场和民用．1987年以来，解析测图仪进入发展的第三个阶段，成为面向数字测图、地形数据库和地理信息系统的摄影测量数据采集系统．中国于20世纪80年代研制和成批生产解析测图仪，主要有国家测绘局测绘科学研究所和无锡测绘仪器厂联合生产的JX-1和JX-3型，西安测绘研究所和1001厂联合生产的APS-1型和APS-2型解析测图仪．

正射投影仪(orthoprojector)　实现微分纠正以获得正射相片图的仪器．除了利用电子计算机对数字影像进行微分纠正外，目前主要利用计算机控制的光学投影晒像的正射投影仪．这是一种间接的函数投影的割面方式的微分纠正仪器．当对正射相片上一个个小缝隙晒像时，仪器在计算机控制下自动将原始相片上相应的影像通过平移、旋转和缩放实现投影晒像，并严格保证缝隙两端点所对应的物像满足中心投影和正射投影的对应关系．

微分纠正经历了很长的历史发展过程．目前市场上出售的主要有瑞士威特厂的Avioplan OR-1(1976年问世)、联邦德国欧波同厂的解析正射投影仪Orthocomp Z-2(1980年问世)和中国刘先琳设计的ZS-1数控正射投影仪(1985年问世)．

数控正射投影仪的硬件包括投影晒像设备、控制驱动部件和电子计算机(或过程计算机)；软件包括六个标准程序，它们是参数设置程序、定向程序、密度测定程序、扫描程序、晒印注记程序和列表打印程序．为了保证正射投影仪的正常作业，需要提供相应的相片断面数据或数字高程模型．此外正射投影

仪亦可用作不同投影方式图象间的变化.

数字摄影测量系统(digital photogrammetry system) 简称DPS.是利用人工和自动化技术,由数字影像产生摄影测量产品,或为地形数据库、地理信息系统(GIS)以及计算机辅助设计(CAD)和计算机辅助制造(CAM)系统采集数据的计算机系统.它由标准的计算机硬件和专用的摄影测量软件构成.

系统硬件包括数字影像扫描输入设备(实现模数转换)、数字摄影测量工作站(实现语义和非语义信息提取)和输出设备(实现成果可视化和数模转换).三大部分可以联机,也可以脱机作业.鉴于处理摄影测量立体像对的需要,系统一般装有各种形式的立体观察装置.

系统软件用来由数字影像提取语义和非语义信息,并尽可能地提高自动化和智能化程度.利用机载、星载GPS技术和高精度多片影像匹配确定相片控制点和连结点,借助空中三角测量,联合平差软件,可以全自动测定点位.利用各种数字影像匹配方法自动建立数字表面模型和生成数字正射影像的软件已推向市场,产品有美国海拉瓦(Helava,V.)的HAI-750、德国阿克曼(Ackermann,W.)教授的Matching-T以及中国武汉测绘科技大学的VirtuoZo.

摄影测量中的地物测绘属于语义信息的提取,可采用人机交互的半自动化方法,也可采用基于影像分析和影像理解的全自动化方法.

数字摄影测量系统的出现,从根本上改变了长期以来摄影测量的生产格局,并使得摄影测量与遥感图象处理及计算机视觉等高技术结合起来.

激光测卫(测月)系统(earth-to satellite (moon) laser ranging system) 以激光为脉冲信号源,测定地面站到人造卫星(月球)表面反射镜之间精确距离的全套仪器设备.通常是在地面站安装激光测距及其附属设备,在卫星(月球)表面安装反射镜.测量时测距仪对准卫星(月球),激光器发射激光脉冲,其中有很微小一部分脉冲能量被送到测时装置,作为计时的开启脉冲,而绝大部分脉冲能量则到达卫星(月球),再经反射镜反射回来被测距仪接收,同时一部分转换为电脉冲被送到测时装置作为关闭脉冲.计数器记下激光脉冲往返的时间,便可利用光速国际值算出所测距离.测距仪应包括激光的发射和接收部分、计时部分、跟踪瞄准卫星(月球)部分、计算机自控和数据处理部分.

用于测卫的测距仪有固定式和流动式两种.前者体积较大,安装在固定台站;后者可做得较小,便于安装在车辆上使用.能供激光测距的卫星表面需嵌有反射镜(因卫星自身旋转难以保证固定的一面朝向地球,需在其表面嵌满反射镜).这种卫星多是专用卫星,小而重.例如美国1976年发射的莱杰奥斯(LAGEOS)卫星,直径仅60cm,球形,重411kg,表面嵌有426块棱镜,离地面6000km,寿命很长,如无意外,估计可至百万年.

国际上研究激光测卫技术的国家先后有美国、法国、德国、俄罗斯、英国、中国等,测距精度也不断提高,从20世纪60年代的米级发展到至今的1～2cm.与所测距离相比,相对误差为10^{-9}～10^{-8}.20世纪末各国布设在全球各地的跟踪站所用的激光测卫仪估计有上百台.中国也有好几个台站,均使用本国研制的测距仪,其中国家地震局武汉地震所研制的DIR-Ⅲcm级卫星激光测距仪达到了国际水平.

研制激光测月系统的国家主要是美国和苏联,两国还先后在月球表面安置了5块反射镜,并取得了许多观测资料.月球可以固定一面朝向地球,安置少量反射镜即够.但对激光测距仪的功率、性能等技术要求则要高得多.目前测距精度约为±10cm,相对误差为10^{-9}.

激光测卫成果可用于地球物理学、大地测量学和天文学研究课题,如大陆漂移和地壳形变、精密测定卫星轨道和地面点地心坐标、地球自转和地极变化,还可用于改善月球星历、推算地球引力参数*GM*等研究课题.

GPS **测地系统**(GPS geodetic system) 利用GPS卫星信号测定地面点的三维或四维坐标的设备和数据处理系统,包括GPS天线、GPS信号接收和处理器(统称为GPS接收机)以及测后数据处理系统.

依据利用GPS卫星信号资料的不同,为实施不同方法的定位,GPS接收机的设计也各不相同.其方法主要有伪距测量和载波相位测量定位.但是,GPS接收机往往不是单纯地为一种定位方法而设计,即一台GPS接收机可以实施两种方法的定位.

依据GPS接收机在实施定位作业时相对于地面处于静止或运动状态的不同,或者说欲测定点的位置在所建立的数学模型中能否视作常数,又可将GPS测地方法分为静态定位和动态定位,以组成相应的GPS测地系统.测后数据处理系统包括GPS观测值的预处理和基线向量计算、GPS基线向量网的平差、定位成果的坐标系转换以及GPS动态定位数据处理等软件系列.

GPS **导航系统**(GPS navigation system) 利用GPS卫星信号为地面、水面上下和空间的各类运行器实施动态定位的系统.动态定位可依据应用目的和精度要求归结为:单点动态定位、实时差分动态定位和测后处理差分动态定位.据此,可建立如下不同的GPS导航系统:

1.GPS单机导航系统.安置于运行器上的GPS

接收机自主测定其实时位置，从而描绘出运行器的运行轨迹，实现导航.

2. 差分 GPS 导航系统. 该系统又可分为实时差分导航系统和测后处理差分导航系统. 由安置于运行器上的 GPS 接收机和安置于固定基准台站上的 GPS 接收机组成. 在实施实时差分导航时，要求建立运行器与基准站之间的实时数据传输系统，以实时联合处理两 GPS 接收机测得的数据，实现实时导航. 而测后处理差分导航系统用于对运行器和基准站 GPS 接收机所采集的定位数据进行测后联合处理，计算出运行器在不同时刻的位置，所以无需上述实时数据传输系统.

3. 为改善 GPS 卫星的全球覆盖不完善的缺陷，可与惯性导航组成 GPS 惯性综合导航系统. 该系统系 GPS 接收机与惯性导航系统的一种或松散、或紧密的综合. 依据不同的导航要求，可选择不同的综合水平.

立体摄影机(stereocamera) 同步摄影得到立体像对的摄影机. 两架内方位元素相同的摄影机装配在一个基线架两端，基线值为已知(一般是定值，亦有可变的)，两摄影光轴平行，由同步器控制同时曝光，摄得立体像对. 广泛用于近景摄影测量、建筑摄影测量、工业摄影测量及交通事故和刑事案件现场摄影等.

重力仪(gravimeter) 亦指相对重力仪器. 即用于测量某点的重力相对于另一(重力值已知)点的重力值之差的仪器. 它是利用在不同环境中仪器的物理参数的变化，进而推导出重力值的变化. 目前广泛应用的重力仪属弹性重力仪，它是用重力作用于弹簧的伸缩性来测量重力变化的. 其中精度较高的是美国产的拉科斯特-隆贝格重力仪(LCR 重力仪)，其误差约为 0.1mgal($1\text{mgal}=10^{-8}\text{m.s}^{-2}$).

绝对重力仪(absolute gravimeter) 一种测量重力值大小的仪器. 基本原理为：在重力场中测定质体在不同状态下的运动轨迹，利用它与重力的关系确定相应的重力值. 仪器可分为振摆式和自由落体式两种类型. 振摆式是利用测定固定摆长的摆动周期来求出绝对重力值；自由落体式是测量物体在真空中下落的时间与距离的关系以得到重力值. 后者还可分为自由落体式和对称上抛下落式. 由于时间与距离测量精度的提高，绝对重力仪测量的绝对重力值的精度已为几个微伽($1\mu\text{gal}=10^{-6}\text{cm}\cdot\text{s}^{-2}$). 目前世界上的绝对重力仪大多数是自由落体式.

超导重力仪(superconductor gravimeter) 一种以超导球体和超导线圈等构成的测量重力的仪器. 在低温(≤10 K)状态下，超导线圈在电流作用下产生感应磁场，磁力使得超导球处于漂浮的零位状态，在重力作用下球体偏离零位，用电容传感器测量超导球偏离零位的大小，由此达到测量重力的目的. 超导重力仪由制冷系统、悬浮系统和记录与转换系统共同构成. 该仪器最大优点是近乎没有零点漂移.

重力梯度仪(gradiometer) 测量重力梯度的仪器. 它可给出测点的重力位的所有二阶导数——重力梯度张量或者它们的线性组合. 它的原理是：测量两个或多个实验物体在非均匀重力场中的运动状态，并换算为重力梯度张量. 重力梯度仪一般分为静态和动态两种类型：静态仪器以扭称为代表，多数采用高精度重力仪改制而成；动态仪器是在运动平台上进行的梯度测量，它具有效率高的特点. 此外，它们还可分为地面、航空、卫星重力梯度仪.

测量船(survey ship) 专门执行海洋、江河、湖泊测量调查任务的船舶. 根据不同的任务，在船上安装各种必要的专用仪器和工具，以进行水深测量、底质探测、扫海和海洋大地测量、重力测量、磁力测量、水下工程施工测量、水文观测以及搜集编制海图所需的各种资料等工作. 大型测量船一般执行远海测量调查任务，以综合性测量为主；中型测量船一般执行近海测量任务；小型测量船主要进行沿岸、江河、湖泊水域的测量任务. 此外，还有执行某项专门测量任务且配以专门测量仪器、工具的测量船，如扫海测量船、地磁测量船、海洋调查船等.

多谱段相机(multispectral camera) 多谱段摄影遥感专用相机系统. 它采用适当的感光材料和滤光片，将从同一景物辐射及反射来的光，按不同波长分成若干段，分别以同样大小的比例尺和幅面，同时记录下来. 用这种相机拍摄的一组相片，是同一景物不同光谱段内的辐射像. 多谱段相机有三种类型：

1. 多相机型. 用几架照相机组装的，每架相机配有不同滤光片和胶片组合.

2. 多镜头型. 用一架照相机把几个镜头和滤光片配合起来，在一张胶片上拍摄多种谱段的黑白相片.

3. 光束分离型. 来自地物的光线通过一个镜头射到相机内的二向色分光镜分成几个谱段，拍出各谱段的黑白相片.

多谱段相机要满足下述条件：

1. 镜头光轴的平行性. 保证以同一视角拍摄到的几幅图象是同一地区.

2. 摄影的同步性. 需调整各谱段曝光时间的差异.

3. 相片比例尺的一致性. 需校正因波长不同而引起的光的折射率差异.

4. 图象的重合. 彩色合成时要进行图象的套合.

5. 四周渐晕的消除. 使影像四周部分不发暗，以获得均匀的像质.

6. 减少镜头畸变，保证成像质量.

现在通常使用的是四谱段相机，选用的是红、绿、蓝和近红外四个谱段.

返束光导管摄像机(return beam vidicon camera) 简称 RBVC. 用返束光导摄像管(RBV)作为接收器件的太空电视摄像机. 它由光学系统、快门、滤光片、返束光导摄像管及电子控制设备等部分组成. 其特点是：由电子枪发射的电子束扫描来接收地面的反射光信息，经传输、处理后在荧光屏上显像或记录在磁带上. 光学系统在快门开启时将景物成像于摄像管的靶面上，滤光片使靶面接收的是预先规定的光谱辐射. 曝光时间可以选择. 曝光后，图象在摄像管上作短暂存储，然后由电子束扫描，产生串联的视频信号. 整个工作过程可分为清除、准备、曝光和读出四个步骤. 它有很高的分辨率，现已达到 200 电视线/毫米，约为日常生活中的电视分辨率的 7 倍. 美国在早期的陆地卫星上装有三台独立的返束光导管摄像机. 摄像时，三个光谱段的摄像镜头对准同一目标，这样由三个光谱段可合成彩色摄像图象.

多谱段扫描仪(multispectral scanner) 简称 MSS. 陆地卫星系列上的一种主要遥感器. 由望远系统、扫描镜、滤光片、探测器和电子系统等组成. 光谱范围为 0.5～1.1μm，分四个谱段：4 谱段为 0.5～0.6μm；5 谱段为 0.6～0.7μm；6 谱段为 0.7～0.8μm；7 谱段为 0.8～1.1μm. 其 3 号卫星增加热红外谱段，即 8 谱段为 10.4～12.6μm. 4 号和 5 号卫星的 MSS 各谱段的波长范围和序号有所改变. MSS 的总视场角为 11.56°～14.92°，其 1 号和 2 号卫星影像像元相应于地面 57m×79m，3 号卫星 4～7 谱段影像地面分辨率为 82.6m，8 谱段为 248m，地面覆盖范围为 185km×185km(轨道高度为 900km 时). 4～7 谱段使用光电倍增管探测器，8 谱段使用碲镉汞探测器. 数据主要用于地球资源调查及管理，是农作物估产、森林和草场管理、土地利用、土地覆盖分类、自然灾害影响估价、能源的矿产资源探测以及其他许多地球资源调查的主要太空遥感资料来源.

专题成像仪(thematic mapper) 简称 TM. 又称专题制图仪. 一种具有较高空间分辨率的成像式多谱段扫描辐射计. 已用于陆地卫星 4 号和 5 号卫星上. 该遥感器是陆地卫星多谱段扫描仪的发展. 光谱通道由 4 个增至 7 个，其中包括两个短波红外和一个热红外通道. 它主要由成像光学系统、扫描机械、光谱波段分离光学部件、探测器列阵、辐射致冷设备、飞行校准器以及所需的操作和处理电子系统等组成. 光学孔径为 41cm，影像像元地面分辨率为 30m(1～6 谱段)和 120m(7 谱段)，轨道高度为 705km，扫描带宽为 185km. 数据传输速度为 85Mb/s，数字化精度为 256 级，谱段间配准精度为 0.1 瞬时视场，长期扫描稳定度为 0.5 瞬时视场.

成像光谱仪(imaging spectrometer) 20 世纪 80 年代发展起来的一种新型遥感器. 它是在 20 世纪 70 年代问世的线阵列遥感器的基础上，使用了二维探测器面阵，从而可同时获取目标影像和该影像各像元的光谱组成，是一种"谱像合一"的遥感器.

物镜将收集到的来自目标的光能聚焦到一条狭缝上，狭缝确定了地面的横穿轨迹扫描线. 通过狭缝的光线被光谱仪的棱镜或光栅色散，然后重新聚焦到放置有二维探测器列阵的焦平面上. 这样，由光谱仪狭缝所确定的横穿轨迹扫描线上的每个像元在许多邻接的光谱带内同时成像，随着遥感平台的运动，狭缝推扫成像，从而获得一组谱像合一的遥感信息.

成像光谱仪的关键是它的探测器——可见光及红外探测器面阵. 它是将可见光及红外探测器与硅 CCD 多路传输器相配接. 这个结构允许探测器上有相当长的积分时间，然后很快地把信号读到硅 CCD 多路传输器上，由电子线路处理.

成像光谱仪是性能较完备的遥感器，将广泛地用于收集高质量的科学数据，大大提高通过遥感来识别地物目标的能力.

红外辐射计(infrared radiometer) 对物体的红外辐射进行绝对测量的红外遥感器. 从测得的红外辐射强度可推算出物体的温度. 它由光学机械扫描系统、红外元件及致冷器、电子系统、电光转换系统、成像光学扫描系统、记录系统等部分组成. 辐射计随平台沿飞行方向前进，扫描镜同时使辐射计的瞬时视场垂直于飞行方向进行行扫描. 地物红外辐射经光学系统会聚在红外元件上，红外元件将辐射信号转换成电信号输出，送到电转换系统、成像光学扫描系统或记录系统. 为了能进行绝对测量，仪器内装有 1～2 个已知的参考黑体源，并在地面严格定标. 常用的红外元件有锑化铟、碲镉汞和热敏电阻等. 获得的信息可由电光转换装置给出图象，也可经计算机处理后得出目标温度. 图象式红外辐射计一般都是扫描成像，它不仅可以获取目标的热图象，还可测温. 工作谱段一般在 3～5μm 和 8～14μm 之间. 非成像式辐射计可分为扫描式与非扫描式两种，主要用于获取目标的辐射信息.

侧视雷达(side-looking radar) 简称 SLR. 一种工作于微波波段的机载成像雷达. 用以获取二维或三维微波图象，主要由脉冲发生器、微波发射机、天线收发开关、侧视天线、接收机、阴极射线管和胶片记录器等组成. 机载侧视天线以一定的倾角俯视地面，其成像地区为飞机一侧或二侧且与航迹相平行的带状地区. 侧视雷达具有快速覆盖和全天候工作能力，其图有明显的阴影，能突出地形特征细微变

化，可用于勘探各种地面资源、测绘地形、判别地质结构特征等. 侧视雷达分真实孔径侧视雷达和合成孔径雷达两类.

合成孔径雷达(synthetic aperture radar) 简称 SAR. 一种高方位分辨率的相干成像雷达. 利用相位和振幅信息，是一个准全息系统. 有侧视、斜视、多普勒锐化和聚束测绘等工作方式. 它利用合成天线技术获取良好的方位分辨率，利用脉冲压缩技术获取良好的距离分辨率.

合成孔径雷达又分聚焦和非聚焦两种. 前者方位分辨率好、与目标距离无关、覆盖面积大、测绘速度快，但设备复杂. 后者方位分辨率与波长和距离的平方根成正比，其所形成的天线长度有一个最大的可能值.

这种雷达主要用于遥感、航空测量和侦察.

微波辐射计(microwave radiometer) 能定量测量目标(如地物和大气各成分)的低电平微波辐射的高灵敏度接收装置. 实质上就是一个高灵敏度、高分辨率的微波接收机. 表面辐射率为 $\varepsilon(0\leqslant\varepsilon\leqslant1)$、绝对温度为 $T_0(T_0>0\ \mathrm{K})$的物体在整个电磁波的频谱上都会辐射出电磁波，其频谱与噪声相似，这种辐射称为热辐射. 不同物体具有不同的热辐射频谱. 有些物体辐射连续频谱，有些物体辐射离散频谱. 通过测量和分析其辐射频谱，就可以区分不同物体.

1946 年，狄克(Dicke，R. H.)首先研制成第一台测量微波辐射的装置，称为狄克式辐射计. 现在各种微波辐射计都是在狄克型接收机基础上改进而成的，有零平衡型辐射计、双参考温度辐射计、自动反馈型辐射计、相关型辐射计、扫描型辐射计等.

微波辐射计的主要技术指标是频段和温度分辨率、空间分辨率. 目前机载微波辐射计实测温度分辨率达 0.02K. 星载微波辐射计温度分辨率达 0.2～1K.

微波辐射计分两类：频谱式和连续式. 前者频率窄，工作于微波谐振线上，后者用于遥感具有宽广频谱特性的目标. 微波辐射计在军事侦察、气象学、海洋学和天文学等领域中得到广泛应用.

微波辐射计还可以分为图象型和非图象型. 其中采用扫描天线的扫描微波辐射计就是图象型辐射计，其特点是天线可以对地面目标进行扫描探测，获取地面目标的微波辐射信息，把所获取的信息转换成以灰度等级显示的物体图象. 扫描方式有两类：

1. 电扫描，如雨云 5 号和 6 号气象卫星上的电扫描微波辐射计.

2. 机械扫描，如雨云 7 号和海洋卫星 1 号上的扫描多通道微波辐射计和泰罗斯 N 号上的微波探测器.

电荷耦合器件(charge coupled device) 简称 CCD. 由时钟脉冲电位来产生和控制半导体势阱的变化，实现存储和传递电荷信息的固态电子器件. 实际上是一种用电荷量来表示不同状态的动态移位寄存器. 其主要应用是成像敏感器、存储器和电信号处理器. 它利用硅集成电路的工艺操作，器件体积小、耗电少、操作可靠、具有扫描功能. 用它制作的遥感器中没有光-机扫描装置，是新一代高级固态器件的基本器件. 硅 CCD 成像敏感器对可见光和近红外辐射(0.4～1.1μm)响应，是重要的探测器. 法国的斯波特遥感卫星上，利用硅 CCD 线列(1728 像元)成像器，从太空得到目前最好的商业遥感图象. 目前，硅 CCD 阵列水平线列为 2048 像元，面阵为 512 像元×512 像元.

20 世纪 90 年代，德国的 MOMS-2 调制光电多光谱立体扫描仪采用位于像面上相互平行并垂直于飞行方向的三个线阵列 CCD 传感器，通过向前、向下和向后扫描，可同时获得三条 CCD 数字影像条带. 最高分辨率可达到地面 5m，且可实时地获得所摄地面的立体影像.

遥感平台(platform for remote sensing) 放置遥感器的运载器具. 按高度大体可分地面平台、空中平台、太空平台三大类. 地面平台包括三角架、遥感塔、遥感车(或船)、建筑物的顶部等，主要用于在近距离上测量地物波谱和摄取供试验研究用的地物细节影像；空中平台包括在大气层内飞行的各类飞行器，如气球、飞艇、飞机等；太空平台包括超出大气层的飞行器，如各种太空飞行器和探空火箭. 这三类平台在具体的遥感作业中，往往是互为补充、相互配合使用的. 为保证遥感数据的质量，飞机上的遥感器通常装在陀螺稳定平台上；太空飞行器一般都要稳定地对地定向.

地球资源技术卫星(earth resource technology satellite) 简称 ERTS. 现称陆地卫星. 美国的一种利用星载遥感器获取地球表面图象数据进行地球资源调查的卫星即资源卫星. 该卫星于 1970 年开始研制，由"雨云"气象卫星改进而成. 1972 年 7 月 23 日发射第一颗地球资源技术卫星，1975 年改名为陆地卫星，至 20 世纪末共发射了五颗卫星，完成了该计划的全部发射任务.

前三颗卫星进入高约 919km 近圆形太阳同步轨道，运行周期为 103.27min，每天绕地球运行 14 圈，每 18 天覆盖全球一次. 星载遥感器有：

1. 三台返束光导管摄像机，以三个不同谱段每隔 25s 对准同一地面进行同步成像，地面分辨率为 100m.

2. 多谱段扫描仪，在绿、红和近红外区的四个谱段工作，地面分辨率为 80m.

陆地卫星 4 号和 5 号进入高约 700km 的近圆

形太阳同步轨道，倾角为98.25°，每16天覆盖全球一次. 星上除带有与前三颗基本相同的多谱段扫描仪外，还带有一种先进的遥感器——专题成像扫描仪，可在包括可见光、近红外和热红外在内的7个谱段工作，每幅图象的信息量达300MB，是多谱段扫描仪图象的近10倍，其数据传输速率为85Mb/s，地面分辨率为30m(其中第六谱段的地面分辨率为120m). 专题成像扫描仪的图象数据通过X波段发射机向地面站实时发送数据，也能用Ku波段通过美国宇航局在地球同步轨道上的跟踪和数据中继卫星把数据传送到美国境内的白沙地面站.

卫星上还带有数据收集装置，接收分布在地面上的无人管理数据收集台站发送的资料(水质、流速、温度、雨量等)，并转发给地面接收站. 地球资源技术卫星获得地面图象不受国界限制，数量多，信息丰富，已广泛用于农业、林业、地质、地理、水文、海洋和环境监测等方面.

资源卫星(resource satellite)　见“地球资源技术卫星”.

陆地卫星(landsat)　见“地球资源技术卫星”.

海洋卫星(seasat)　主要用于海洋研究的卫星，如测量海洋的波浪、潮汐、海流、海面温度和海洋污染等. 1978年6月26日美国发射了海洋卫星1号. 它是一颗高度为800km的极轨人造地球卫星.

斯波特卫星(satellite probatoire d' observation de la terrestre)　简称SPOT. 法国的地球观测卫星. 1978年2月正式开始研制，由阿里安火箭送入822km高的太阳同步轨道，重复周期26天. 卫星重1830kg，采用三轴稳定，工作寿命三年. 第一颗斯波特试验卫星1986年发射成功. 主要目的是土地利用调查、植物长势监测、土壤水含量评价、海洋与海岸相互作用研究等方面.

斯波特1号、2号卫星携带两台相同的高分辨率CCD相机，能在可见光和近红外谱段工作，电子自扫描成像. 每台相机扫描宽度为60km，地面分辨率分别为20m(3个谱段)及10m(全色). 斯波特2号于1990年1月发射入轨. 到20世纪90年代中期，发射经过改进的斯波特3号和4号，星载相机增加了热红外谱段，以扩大使用范围；影像分辨率提高为5m；卫星工作寿命提高到4年.

斯波特卫星在设计上有很多特点，如采用多功能平台，使卫星的大部分硬件可重复使用；星载遥感器的地面分辨率较高，并具备立体制图能力；相机装有可变指向反射镜，能在偏离星下点±27°(最大可达30°)范围内观测任何区域，平均两天就有可能对任意给定地区拍摄一次，缩短了重复观测的时间等.

航天飞机(space shuttle)　往返于地面和近地轨道之间的可重复使用的太空飞行器. 它同时具备火箭、飞机和飞船的特点，能在大气层和大气层以外的太空飞行，使用和维修均比一般飞船简便. 可用于运输人员和物资，实施太空发射、维修、装配以及执行太空遥感和军事侦察任务. 美国1981年4月12日发射了世界第一架航天飞机哥伦比亚号，以后挑战者号、发现号和亚特兰蒂斯号也相继投入使用. 它们曾携带成像雷达、多谱段红外辐射仪和大像幅相机等遥感仪器，对地面、海洋、大气和太空目标进行探测. 1981年11月12日至哥伦比亚号飞行期间，宇航号利用侧视成像雷达拍摄了地质构造图象，用多谱段红外辐射计探测了岩石结构，用一种特制相机拍摄了雷电现象，证实了航天飞机的综合太空探测能力. 1984年10月，挑战者号首次携带一种大像幅相机，分辨率为20m，拍摄了地球陆地面积的10%，对地质研究和地图绘制均具有重要意义.

撰　稿　王昆杰　边馥苓　仲维立　李征航　李德仁
宋载仁　张正禄　张儒杰　金标仁　於宗俦
郝正焕　胡毓钜　晁定波　徐庆荣　徐德宝
郭俊义　黄世德　舒　宁　操华胜　薄志鹏
审　阅　史惠顺　宁津生　李德仁　於宗俦　陶本藻

数 学 符 号 表

数学符号表编写说明

《数学辞海》第一至五卷正文之后，均附有数学符号表，提供读者查阅之用. 本表所收符号比较齐全，除包含“中国数学物理名词委员会”审定的《数学物理符号表》中的全部数学符号外，还收入了国内外数学界已普遍使用的数学符号，总共列入数学符号 1158 个.

一些新兴学科，如小波分析、分形几何、数理语言学、机器证明等，都是 20 世纪中叶以后发展起来的，这些学科的数学符号在国际国内还不统一，《数学辞海》将其收入，仅供读者参考.

本表所收数学符号并非仅限于《数学辞海》的正文，有的符号虽然在本辞书的正文中(如模糊数学中的一些专用数学符号)未曾出现，但由于这些符号已经广泛应用于国内外的教学、科研、工程技术中，因此亦作了适当的搜集，以飨读者.

数学符号表的体例：数学符号表共设五个横栏，依次为符号栏、中文名称栏、英文名称栏、意义或举例栏、备注栏.

数学符号的编排分类：《数学辞海》共六卷，包含数学科学的 100 多个分支学科或专题项目，所涉及的数学符号种类繁多. 为便于读者查找而采取分类编排. 因此，本表将数学符号按学科类型分为以下 7 类：

1. 算术与数论：算术中包括最常用的数学符号，如＋，－，×，÷，＝，≠等，它的应用范围遍及所有分支学科. 数论则包括初等数论、代数数论、解析数论、几何数论等.

2. 逻辑与集合：包括数学基础、形式逻辑、数理逻辑、集合论、公理集合论、序与格等.

3. 几何与拓扑：包括平面几何、立体几何、平面三角、球面三角、解析几何、高等几何、微分几何、凸集几何、距离几何、一般拓扑学、代数拓扑学与流形拓扑学等.

4. 代数学：包括初等代数、高等代数、布尔代数、线性代数与多重线性代数、环与代数、模与同调代数、群及其推广、域与伽罗瓦理论、李群与李代数、范畴论与代数 K 理论、代数几何、奇点理论与突变理论等.

5. 分析学：包括数学分析、实变函数论、复变函数论、多复变与复空间、测度论、泛函分析、变分法、函数逼近论、调和分析、流形上的分析、位势论、凸分析、非标准分析、小波分析、分形几何、常微分方程、偏微分方程、积分方程与函数方程、动力系统、特殊函数等.

6. 概率统计：包括组合学、概率论、随机过程、统计学等.

7. 应用数学：包括计算数学、模糊数学、生物数学、经济数学、数学物理与理论物理、运筹学、系统理论、控制理论、通信与信息理论、测绘学、力学、天文学、数理语言学等.

数学符号表的编排顺序：本表所列数学符号，大体上按它们在《数学辞海》中出现的先后顺序编排. 由于很多数学符号的含义及使用范围比较复杂，若要准确地归入哪一类，实际上是很困难的，因而制订下列编排原则：

1. 多学科共用符号，将其编入最先出现的分支学科中. 例如，运算符号＋，－，×，÷等，是所有学科共用的，就编入本表最前面的学科——算术中.

2. 同形同义的符号，就只在某一分支学科符号表内出现一次. 例如，符号“R”在集合论中表示实数集，而在代数学和分析学中也表示实数集，其意义是相同的，就将符号“R”只列入集合论的符号表，而在代数学和分析学的符号表中不再出现.

3. 同形而不同义的符号，则分别列入相应分支学科. 如“Im”在初等代数中表示复数的虚部，而在集合论和代数学中则表示映射的像，就将其分别列入各个学科的符号表中；又如“k”在应用数学中表示高斯常数，在微分几何中表示曲率，而在特殊函数中则表示贝克函数，这样便分别将其列入应用数学、微分几何、特殊函数的符号表中.

4. 异形同义的符号，首先将《数学物理符号表》中核定的符号列入符号栏，而将其异形符号列入备注栏，如几何中将 Rt∠列于符号栏，而将曾用符号 rt∠和 R∠列入备注栏；其次，凡目前国际国内用法尚未统一的异形同义符号，如代数中的“A^T”，“A'”都表示矩阵 A 的转置矩阵，则一同列于符号栏.

5. 过去用过，而现在少用或不用的数学符号，本表将其列入备注栏，以利读者阅读古旧数学资料时参考.

算术和数论(Arithmetic & Number theory)

符号	中文名称	英文名称	意义或举例	备注
+	加号;正号	plus ;positive	例如,+2即正2;$a+b$即a与b相加	正号常可略去不写
−	减号;负号	minus ;negative	例如,−1即负1;$a-b$即a与b的差	
±	正或负;加或减	positive or negative;plus or minus	例如,±2,即正2或负2;$a\pm b$即a加或减b	
∓	负或正;减或加	negative or positive;minus or plus	例如,∓2即负2或正2;$a\mp b$即a减或加b	
×,·	乘号	multiple sign	例如,2×3即2乘3;$a\cdot b$即a乘b	乘号在括号前或字母间常可略去
÷,—,/	除号;分数(式)线	sign of division, fraction stroke	$a\div b$, $\frac{a}{b}$, a/b,即a除以b,b分之a	
:	比	ration	$a:b$即a比b	
|	整除	exact division	$a\|b$即整数a整除整数b	
∤	不能整除	nonaliquot	$a\nmid b$即整数a不能整除整数b	
∥	限界整除	bound exact division	$a^k\parallel b$即a^k能整除b,但a^{k+1}不能整除b	$a^k\|b$,且$a^{k+1}\nmid b$
[,…,]	最小公倍数	least common multiple	$[a_1,a_2,\cdots,a_n]$表示整数$a_1,a_2,\cdots,a_n$的最小公倍数	亦可用LCM表示
(,…,)	最大公约数	greatest common divisor	$(a_1,a_2,\cdots,a_n)$表示整数$a_1,a_2,\cdots,a_n$的最大公约数	亦可用GCD表示
a^n	a的n次方(幂)	a to the power n	例如,5^4即5的4次方(幂)	当$n=2,3$时,分别称平方、立方
$\sqrt{\ }$	平方根号	square root sign	$\sqrt{a}$ 即a开平方	
$\sqrt[n]{\ }$	n次根号	n-th root sign	$\sqrt[n]{a}$ $(n\geqslant 2)$即a开n次方	当$n=3$时,称a开立方
| |	绝对值;模	absolute value;modules	$\|a\|$表示a的绝对值或模	亦可用abs a表示
=	等号	equal sign	$2+3=5$	
≠	不等号	inequality sign	$2+3\neq 4$	
≡	恒等号	identity symbol	$a\equiv b$即a恒等于b	
<	小于	less than	$a<b$即a小于b	
>	大于	greater than	$a>b$即a大于b	
⋛	大于或小于	greater than or less than	$a\gtrless b$即$a>b$或$a<b$	
⋚	小于或大于	less than or greater than	$a\lessgtr b$即$a<b$或$a>b$	
⩽	小于或等于;不大于	less than or equal to	$a\leqslant b$即a小于或等于b,或a不大于b	一般不用符号"≦"
⩾	大于或等于;不小于	greater than or equal to	$a\geqslant b$即a大于或等于b,或a不小于b	一般不用符号"≧"
≪	远小于	much less than	$a\ll b$即a远小于b	
≫	远大于	much greater than	$a\gg b$即a远大于b	
≈	约等于	approximately equal	$a\approx b$即a约等于b	曾用≐,现已不用
≙	相当于	equivalent to	1 cm≙10 km表示图上1 cm相当于实际距离10 km	曾用⇔,现已不用
∝	成正比	is direct ratio to	$a\propto b$表示a与b成正比	
~	数值范围	numerical range	例如,5~10即由5至10	现已不用"—"
.	小数点	decimal point	例如,8.59即8又100分之59	小数点记于个位数字后的下足
· ·	循环小数	recurring decimal	$2.4\dot{2}3\dot{1}$即2.423 123 123 1…	记于循环节的首末位数字上方

符号	中文名称	英文名称	意义或举例	备注
%	百分号	sign of percent	例如，5%即百分之五，亦即 5/100	
‰	千分号	sign of permillage	例如，5‰即千分之五，亦即 5/1000	
( )	圆括号	parenthesis	例如，5−(2+1)	亦称小括号
[ ]	方括号	square brackets	例如，3[5−(2+1)]	亦称中括号
{ }	花括号	brace	例如，2{3[5−(2+1)]−2}	亦称大括号
——	括线	vinculum	例如，$(\overline{8-2}\times 3)\div 2$，以 8−2 的差乘 3…	相当于小括号
∞	无穷大	infinity	$\lim\limits_{x\to 0}\frac{1}{x}=\infty$ 即函数 $\frac{1}{x}$ 当 x 趋近于 0 时无限地增大	亦称无限或无限大
$a\overset{\text{def}}{=}b$	a 以 b 为定义	a is definition equal to b	例如，$a\overset{\text{def}}{=}b^n$ 即用 b^n 代表 a	亦可用 $a\overset{\text{d}}{=}b$ 或 $a:=b$ 表示
d	公差	common difference	等差数列任相邻两项之差(后项减前项)均相等，这个共同的差 d 称为此数列的公差	
q	公比	common ratio	等比数列任相邻两项之比(后项比前项)均相等，这个共同的比 q 称为此数列的公比	
S_n	数列前 n 项和	sum of the first n terms	例如，等差数列 $a,a+d,\cdots,a+(n-1)d,\cdots$，前 n 项之和 $S_n=na+\frac{n(n-1)}{2}d$	
Δ	判别式	discriminant	例如，实系数一元二次方程 $ax^2+bx+c=0(a\neq 0)$ 的判别式 $\Delta=b^2-4ac$	利用 Δ 可判别该方程根的状况
$E(x)$，$[x]$	整数部分记号	symbol of integral part	表示不超过 x 的最大整数. 例如，$[1.2]=1$，$[-1.2]=-2$	亦记为 ent(x)，来自法文 entier
$\{x\}$	小数部分记号	symbol of decimal part	$\{x\}$ 只能是 0 或正的纯小数，它满足：$0\leqslant\{x\}<1$，例如，$\{1.2\}=0.2$，$\{-1.2\}=0.8$	亦称分数部分记号，亦记为 $\langle x\rangle$
$\sum\limits_{n\leqslant x}$	整数求和号	sign of integers summation	对不超过 x 的正整数 n 求和. 例如，$\sum\limits_{n\leqslant 6}n=1+2+3+4+5+6=21$	
$\sum\limits_{n<x}$	整数求和号	sign of integers summation	对小于 x 的正整数 n 求和. 例如，$\sum\limits_{n<6}n=1+2+3+4+5=15$	
$\sum\limits_{p\leqslant x}$	素数求和号	sign of prime number summation	对不超过 x 的素数 p 求和. 例如，$\sum\limits_{p\leqslant 7}p=2+3+5+7=17$	
$\sum\limits_{p<x}$	素数求和号	sign of prime number summation	对小于 x 的素数 p 求和. 例如，$\sum\limits_{p<7}p=2+3+5=10$	
$\sum\limits_{d\mid n}$	除数求和号	sign of divisor summation	对 n 的所有不同因子 d 求和. 例如，$\sum\limits_{d\mid 6}d=1+2+3+6=12$	
$\prod\limits_{d\mid n}$	除数求积号	sign of divisor mensuration	对 n 的所有不同因子 d 求积. 例如，$\prod\limits_{d\mid 6}=1\cdot 2\cdot 3\cdot 6=36$	
$\sum\limits_{p\mid n}$	素除数求和号	sign of prime divisor summation	对 n 的所有不同素因子 p 求和. 例如，$\sum\limits_{p\mid 6}=2+3=5$	
$\prod\limits_{p\mid n}$	素除数求积号	sign of prime divisor mensuration	对 n 的所有不同素因子 p 求积. 例如，$\prod\limits_{p\mid 6}=2\cdot 3=6$	
$\sum\limits_{i=1}^{n}$	总和号	sign of grand sum	求对 x_i 从 x_1 连加到 x_n 的总和，即 $\sum\limits_{i=1}^{n}x_i=x_1+x_2+\cdots+x_n$	
$\prod\limits_{i=1}^{n}$	连乘号	sign of continued product	求对 x_i 从 x_1 连乘到 x_n 的积，即 $\prod\limits_{i=1}^{n}x_i=x_1x_2\cdots x_n$	
$a\equiv b(\bmod n)$	模 n 同余	congruence modulo-n	用 n 除 a 及 b 所得余数相同	
$a\not\equiv b(\bmod n)$	模 n 不同余	non-congruence modulo-n	用 n 除 a 及 b 所得余数不同	
$\overline{\overline{\equiv}}$	恒等同余	identity congruence	$f(x)\overline{\overline{\equiv}}g(x)\ (\bmod p)$，即整系数多项式 f 与 g 的对应系数均模 p 同余	亦可记为 $f(x)\equiv_x g(x)(\bmod p)$
$\overline{\overline{\not\equiv}}$	不恒等同余	non-identity congruence	$f(x)\overline{\overline{\not\equiv}}g(x)\ (\bmod p)$，即 $f(x)$ 与 $g(x)$ 的对应系数均模 p 不同余的	亦可记为 $f(x)\not\equiv_x g(x)(\bmod p)$

符号	中文名称	英文名称	意义或举例	备注
$a^{-1}(\bmod n)$	模 n 的逆	inverse of modulo-n	与 a 相乘后用 n 除余数是 1 的整数. 例如, $2^{-1}(\bmod 5)=3,\ 3^{-1}(\bmod 4)=3$	这是一个同余类
$r \bmod n$	模 n 的同余类	congruence class of modulo-n	包含 r 的模 n 的同余类. 例如, $2(\bmod 5)=\{\cdots,-8,-3,2,7,12,\cdots\}$	亦称剩余类
$\mathbf{Z}_n$	剩余类环	residue class ring	模 n 的全体剩余类对类的加法和乘法组成的环	
$\left(\frac{a}{p}\right)$	勒让德符号	Legendre's symbol	$\left(\frac{a}{p}\right)=\begin{cases}1, & p\nmid a,\text{且 } a \text{ 是二次剩余}(\bmod p)\\ -1, & p\nmid a,\text{且 } a \text{ 是二次非剩余}(\bmod p)\\ 0, & p\mid a\end{cases}$	p 为奇素数, a 为整数
$\left(\frac{a}{m}\right)$	雅可比符号	Jacobi's symbol	$\left(\frac{a}{m}\right)=\prod_{i=1}^{k}\left(\frac{a}{p_i}\right)$ ($m=p_1p_2\cdots p_k$, p_i 为素数, $(m,a)=1$)	当 m 为奇素数时即勒让德符号
$\left(\frac{d}{m}\right)$	克罗内克符号	Kronecker's symbol	$\left(\frac{d}{m}\right)=\prod_{r=1}^{v}\left(\frac{d}{p_r}\right)$ (d 为非平方数, p_r 为素数, $m=\prod_{r=1}^{v}p_r$)	
$d(n)$	除数函数	divisor function	$d(n)$ 表示 n 的正因子的个数. 例如, $d(12)=6$	亦可用 $\tau(n)$ 或 $T(n)$ 表示
$d_k(n)$	广义除数函数	generalized divisor function	$d_k(n)=\sum_{n_1n_2\cdots n_k=n}1=\sum_{m\mid n}d_{k-1}(m)$	
$\sigma(n)$	除数和	sum of divisor	表示正整数 n 的所有正因数的和. 例如, $\sigma(6)=1+2+3+6=12$	亦可用 $S(n)$ 表示
$\sigma_\lambda(n)$	广义除数和	generalized sum of divisor	$\sigma_\lambda(n)=\sum_{d\mid n}d^\lambda$. 例如, $\sigma_3(4)=1^3+2^3+4^3$	$\sigma_0(n)=d(n)$ 为除数函数; $\sigma_1(n)=\sigma(n)$ 为除数和
$P(n)$	正因数之积	product of positive divisors	$P(n)=\prod_{d\mid n}d$. 例如, $P(6)=1\cdot2\cdot3\cdot6=36$	
$\Phi(n)$	欧拉函数	Euler's function	表示小于正整数 n, 且与 n 互素的正整数的个数. 例如, $\Phi(6)=2$	亦可记为 $\varphi(n)$
$\mu(n)$	默比乌斯函数	Möbius function	$\mu(n)=\begin{cases}1, & \text{当 } n=1 \text{ 时},\\ 0, & \text{当 } n \text{ 能被素数的平方整除时},\\ (-1)^r, & \text{当 } n \text{ 为 } r \text{ 个相异素数之积时}\end{cases}$	
$\Lambda(n)$	曼戈尔特函数	Von Mangoldt function	$\Lambda(n)=\begin{cases}\ln p, & n \text{ 为素数 } p \text{ 的正乘方};\\ 0, & \text{其他}\end{cases}$	
$\Lambda_1(n)$	曼戈尔特函数 I	Von Mangoldt function I	$\Lambda_1(n)=\begin{cases}\frac{1}{m}, & \text{若 } n \text{ 是一素数的 } m(>0) \text{ 次乘方},\\ 0, & \text{其他}\end{cases}$	
$\omega(n)$	相异素因数个数	different prime factor numbers	例如, $\omega(24)=\omega(2^3\cdot3)=1+1=2$, 即 24 有 2 个不同的素因数	
$\Omega(n)$	素因数个数	prime factor numbers	表示正整数 n 的所有素因数的个数. 例如, $\Omega(24)=\Omega(2^3\cdot3)=3+1=4$	
$\lambda(n)$	刘维尔函数	Liouville's function	$\lambda(n)=(-1)^{\Omega(n)}$	
$\pi(x)$	素数个数符号	symbol of the prime numbers	表示不超过正实数 x 的素数个数. 例如, $\pi(10)=4$	
$\chi(n)$	特征函数	characteristic function	对模 m 之一特征 $\chi(n)$ 仅在 $(n,m)=1$ 时有定义, 且 $\chi(1)\neq0$; 若 $a\equiv b(\bmod m)$, 则 $\chi(a)=\chi(b)$; $\chi(ab)=\chi(a)\chi(b)$	若 $(n,m)>1$ 时, 则 $\chi(n)=0$
$p(n)$	整数分拆函数	integral partition function	把正整数 n 分成若干个正整数的和, 称为 n 的一种分拆, 以 $p(n)$ 表示分拆的种数. 例如, $p(4)=5$. 若限定分拆中的加数不超过 r, 则这类分拆数以 $p_r(n)$ 表示	
$U(n)$	奇分拆	odd partition	$U(n)$ 为把 n 分为奇数个互异数之和的分拆数	
$E(n)$	偶分拆	even partition	$E(n)$ 为把 n 分为偶数个互异数之和的分拆数	
$N(\mathfrak{m})$	模 $\mathfrak{m}$ 的矩	moment of module $\mathfrak{m}$	将所有线性型依 $\bmod \mathfrak{m}$ 分类, 则分类的个数称为模 $\mathfrak{m}$ 的矩. 若模 $\mathfrak{m}$ 对应于方阵 A, 则 $N(\mathfrak{m})=\det A$	
$\vartheta(x)$	切比雪夫函数	Chebyshev function	$\vartheta(x)$ 表示对不大于 x 的素数的对数求和	
$\psi(x)$	切比雪夫函数	Chebyshev function	$\psi(x)=\sum_{n\leqslant x}\Lambda(n)=\sum_{p^m\leqslant x}\ln p$, 而 $\Lambda(n)$ 为曼戈尔特函数	

符 号	中文名称	英 文 名 称	意 义 或 举 例	备 注
$\zeta(s)$	黎曼 ζ 函数	Riemann ζ-function	$\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}$,其中 s 为实部大于 1 的复数	
∂°	多项式的次数	degree of a polynomial	$\partial^\circ f=n$,表示多项式 $f(x)$ 的次数为 n	亦可表示成 $\deg f=n$
max( )	最大数	maximum number	$\max(a,b,\cdots,c)$ 即 $a,b,\cdots,c$ 中的最大数	
min( )	最小数	minimum number	$\min(a,b,\cdots,c)$ 即 $a,b,\cdots,c$ 中的最小数	
$\overset{L}{=\!=}$	左结合	left association	$A\overset{L}{=\!=}B$ 表示存在模方阵 U,使 $A=UB$,并称方阵 B 左结合于方阵 A	
$[,\cdots,]$	有限连分数	finite continued fraction	$[a_0,a_1,\cdots,a_N]=a_0+\frac{1}{a_1+}\frac{1}{a_2+}\cdots+\frac{1}{a_N}$,即有理数化成的连分数	无理数化成的连分数为无限连分数
Δ	判别式	discriminant	$\Delta(\alpha_1,\alpha_2,\cdots,\alpha_n)$ 表示 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 的判别式; $\Delta=\Delta(R(\theta))$ 表示代数数域 $R(\theta)$ 的判别式	
ind n	指数	index	如果 $n\equiv g^a(\mathrm{mod}\ m)$,则称 a 为 n 对于模 m 且以 g 为底的指数,记为 $a=\mathrm{ind}_g n$,简记为 ind n	亦可用 $\delta_m(a)$ 表示 a 对模 m 的指数
$x^k\equiv n(\mathrm{mod}\,p)$	k 次剩余	residue of degree-k	$x^k\equiv n(\mathrm{mod}\,p)(p\times n)$ 有解,则 n 称为 p 的 k 次剩余	
$d(A)$	A 的密率	density of A	$d(A)=\inf_{n\geqslant1}\frac{A(n)}{n}$,即集 A 的密率为 $A(n)/n$(一切 $n\geqslant1$)的下确界	$A(n)$ 表示 A 中不大于 n 的正整数的个数
$\delta^*(A)$	A 的渐近密率	asymptotic density of A	$\delta^*(A)=\lim_{n\to\infty}\frac{A(n)}{n}$,即集 A 的渐近密率为 $A(n)/n$ 当 $n\to\infty$ 的极限值	
$\left(\frac{a,b}{m}\right)$	和数符号	sum symbol	设 $m>1$,a,b 都是整数,令 $\left(\frac{a,b}{m}\right)=\sum_x e^{2\pi i\frac{ax+bx'}{m}}\left(x'\equiv\frac{1}{x}(\mathrm{mod}\ m)\right)$, 其中 x 是通过与模 m 简化的剩余系	
$(a,b)=\pm1$	希尔伯特符号	Hilbert symbol	设 k^* 表示域 k 的单位群,又 $a,b\in k^*$,则 $(a,b)=\begin{cases}1, & 若\ z^2-ax^2-by^2=0\ 在\ k^3\ 中有非零解,\\ -1, & 其他情形\end{cases}$	
$\{a,b,c\}$	二元二次型	2-ary quadratic form	用 $\{a,b,c\}$ 表示二元二次型 $ax^2+bxy+cy^2$,其中 a,b,c 为整数	
$g(k)$	小 $g(k)$	small $g(k)$	设 k 为一固定正整数,对任意正整数 n,不定方程 $n=x_1^k+x_2^k+\cdots+x_s^k$ 总有解的最小正整数 s	
$G(k)$	大 $G(k)$	large $G(k)$	设 k 为一固定正整数,对充分大的正整数 n,不定方程 $n=x_1^k+x_2^k+\cdots+x_s^k$ 总有解的最小正整数 s	
$S(\alpha)$	α 的迹	trace of α	设 $R(\theta)$ 为 n 次代数数域,$\alpha^{(1)}=\alpha\in R(\theta)$,$\alpha^{(k)}(k=2,3,\cdots,n)$ 为 α 的共轭数,则 $S(\alpha)=\sum_{k=1}^{n}\alpha^{(k)}$ 称为 α 的迹	
$N(\alpha)$	α 的范数	norm of α	$N(\alpha)=\prod_{k=1}^{n}\alpha^{(k)}$ 为 α 的范数	亦称矩
$N(k)$	等幂和	sum of equal powers	使 $x_1+x_2+\cdots+x_s=y_1+y_2+\cdots+y_s,\cdots,x_1^k+x_2^k+\cdots+x_s^k=y_1^k+y_2^k+\cdots+y_s^k$ 的最小正整数 s 记为 $N(k)$,其中 $y_1,y_2,\cdots,y_s$ 不是 $x_1,x_2,\cdots,x_s$ 的重组	
$M(k)$	强等幂和	strong sum of equal powers	使 $x_1+x_2+\cdots+x_s=y_1+y_2+\cdots+y_s,\cdots,x_1^k+x_2^k+\cdots+x_s^k=y_1^k+y_2^k+\cdots+y_s^k$,并使 $x_1^{k+1}+x_2^{k+1}+\cdots+x_s^{k+1}\neq y_1^{k+1}+y_2^{k+1}+\cdots+y_s^{k+1}$ 的最小正整数 s 用 $M(k)$ 表示	
$S(a,\chi)$	特征和	character sum	$S(a,\chi)=\sum_{n=1}^{m}\chi(n)e^{2\pi ian/m}$	
$S(n,m)$	高斯和	Gauss sum	$S(n,m)=\sum_{x=0}^{m-1}e^{2\pi ix^2n/m}$,其中 $(n,m)=1$	
$F(s)$	狄利克雷级数	Dirichlet series	$F(s)=\sum_{n=1}^{\infty}\frac{f(n)}{n^s}$	亦称 $F(s)$ 为 $f(n)$ 的演成函数
M_p	梅森数	Mersenne number	形如 2^p-1(p 为素数)的素数称为梅森数,记为 M_p.例如,$M_2=3$,$M_3=7$	
F_n	费马数	Fermat number	形如 $2^{2^n}+1$ 的数称为费马数,例如,$F_2=17$	F_5 不是素数

符号	中文名称	英文名称	意义或举例	备注
≡	重模同余式	double module congruence expression	$f(x) \equiv g(x) \pmod{p, \varphi(x)}$ 表示系数以素数 p 为模，又 $\varphi(x)$ 整除 $f(x)-g(x)$，称为重模同余式	亦称重模为双模
$Q(x)$	无平方因子数	number of noninclusion square divisor	不超过 x 的无平方因子数的个数. 例如，$Q(10)=6$	
$V(n)$	同余式的解数	number of solutions of congruence expression	同余式 $x^2 \equiv -1 \pmod n$ 之解数	
$R(x)$	圆内整点数	number of circle lattice point	表示圆 $u^2+v^2 \leqslant x$ 内的整点数	
$F(x)$	朗伯级数	lambert series	$F(x)=\sum_{n=1}^{\infty} f(n)\frac{x^n}{1-x^n}$ 称为朗伯级数	
$[\alpha_1,\cdots,\alpha_q]$	理想数	ideal number	$\alpha_1,\alpha_2,\cdots,\alpha_q$ 为 $R(\vartheta)$ 中之整数，$R(\vartheta)$ 中形如 $\eta_1\alpha_1+\eta_2\alpha_2+\cdots+\eta_q\alpha_q$（$\eta_i$ 为 $R(\vartheta)$ 中之整数）的整数所成之集合为理想数	
$[1]$	单位理想数	unit ideal number	表示单扩域 $R(\vartheta)$ 中全体整数组成之集合	
$\tau(n)$	拉马努金函数	Ramanujan function	表示 cusp 型 $F(s)=(2\pi)^{-12}\triangle(Z)$ 的第 n 个系数. 称 $n\mapsto\tau(n)$ 为拉马努金函数	
$L(s,\chi)$	狄利克雷级数	Dirichlet series	表示狄利克雷级数 $\sum_{n=1}^{\infty}\chi(m)n^{-s}$，其中 $m\geqslant 1$ 为整数，χ 为 mod m 特征	
$G_k(\Gamma)$	艾森斯坦级数	Eisenstein series	设 Γ 是 C 格，则称 $G_k(\Gamma)=\sum_{\gamma\in\Gamma}'\frac{1}{\gamma^{2k}}$ 为指标是 k 的艾森斯坦级数，其中 $\sum'$ 表示对 Γ 的非零元素求和	
$\theta_\Gamma(Z)$	塞他函数	theta function	$\theta_\Gamma(Z)=\sum_{x\in\Gamma}e^{\pi i Z(x\cdot x)}$ 称为二次模 Γ 的塞他函数	

逻辑与集合(Logic & Sets)

符号	中文名称	英文名称	意义或举例	备注
$\forall$	全称量词	universal quantifier	$\forall x\in A$，$p(x)$，表示命题 $p(x)$ 对于每一个属于 A 的 x 为真	亦可简记为 $\forall x$，$p(x)$
$\exists$	存在量词	existential quantifier	$\exists x\in A$，$p(x)$，表示存在 A 中的元素 x 使 $p(x)$ 为真	$\exists^1$（或 $\exists!$）表示存在一个且只有一个元素使 $p(x)$ 为真
$\wedge$	合取符号	conjunction sign	$p\wedge q$ 即 p 和 q	
$\vee$	析取符号	disjunction sign	$p\vee q$ 即 p 或 q	
$\neg$	否定符号	negation sign	$\neg p$ 即 p 的否定，非 p	
$\rightarrow,\Rightarrow$	推断符号	implication sign	$p\rightarrow q$，$p\Rightarrow q$ 表示：若 p 则 q，p 蕴含 q	亦可用 $q\leftarrow p$，$q\Leftarrow p$
$\leftrightarrow,\Leftrightarrow$	等价符号	equivalence sign	$p\leftrightarrow q$，$p\Leftrightarrow q$ 表示 $p\Rightarrow q$，且 $q\Rightarrow p$，即 p 等价于 q	亦称充分必要条件
$\models$	真值符号	truth sign	$\models A\rightarrow B$ 表示由命题 A 推出命题 B 为真	
$\Dashv\!\models$	可逆真值符号	invertible truth sign	$A \Dashv\!\models B$（或 $\models A\leftrightarrow B$）表示 $A\models B$，且 $B\models A$，意即 A 真则 B 真，且 B 真则 A 真	亦即 A，B 具有相同的真值
$\vdash$	断定符号	predicative sign	$p\vdash q$ 表示 q 随 p 来，p 是或从一公理而来，或 p 是同语反复	
$\in$	属于	belongs to	$x\in A$ 表示 x 属于 A，即 x 是集 A 的一个元（素）	集合 A 可简称为集 A
$\not\ni$	不包含	noninclusion	$A\not\ni x$ 表示集合 A 不包含元素 x	
$\overline{\in}$，$\notin$	不属于	nonmembership	$y\notin A$，$y\overline{\in}A$ 表示 y 不属于 A，y 不是集 A 的一个元（素）	亦可记为 $A\overline{\ni}y$，或 $A\not\ni y$
$\{,\cdots,\}$	集合号	sign of set	$\{x_1,x_2,\cdots,x_n\}$ 表示由诸元素 $x_1,x_2,\cdots,x_n$ 构成的集	亦可用 $\{x_i,\ i\in I\}$，这里 I 表示指标集
$\{\mid\}$	集合号	sign of set	$\{x\in A\mid p(x)\}$ 即使命题 $p(x)$ 为真的 A 中诸元（素）组成的集	亦可用 $\{x\in A:p(x)\}$ 表示集

数 学 符 号 表

符号	中文名称	英文名称	意义或举例	备注
$\varnothing$	空集	the empty set	$\varnothing$表示没有元(素)的集	$\varnothing$是丹麦文字母，读“欧”
N	非负整数集	nonnegative integers set	N=\{0,1,2,…\}	N_+=\{1,2,3,…\}
Z	整数集	integers set	Z=\{…,−2,−1,0,1,2,…\}	Z_+表示正整数集合
Q	有理数集	rational numbers set	由全体有理数组成的集合	Q_+表示正有理数集合
R	实数集	real numbers set	由全体实数组成的集合	R^n表示n维实空间
C	复数集	complex numbers set	由全体复数组成的集合	C^n表示n维复空间
R^+	正实数集	positive real numbers set	由全体正实数组成的集合	R^-表示负实数集
R^*	扩张的实数集	expanding system of the real numbers	把两个理想点$+\infty$，$-\infty$加进实数系所得的集	亦称扩张的实数系
$\subsetneqq$	真包含于	proper inclusion	$B\subsetneqq A$表示A的子集B真包含于A	亦可用$\subset$表示
$\subseteq$	包含于	inclusion	$B\subseteq A$表示B是A的子集，即B的每一个元素均属于A	
$\not\subset$	不包含于	noninclusion	$C\not\subset A$表示C不是A的子集	亦可用$\nsubseteq$表示
$\supsetneqq$	真包含	proper inclusion	$A\supsetneqq B$表示A真包含B	
$\supseteq$	包含	inclusion	$A\supseteq B$表示B是A的子集	亦可用$\supset$表示
$\not\supset$	不包含	noninclusion	$A\not\supset C$表示A不包含C	亦可用$\nsupseteq$表示
$\cup$	并集，和集	union	$A\cup B=\{x\mid x\in A\vee x\in B\}$，称为$A$与$B$的并集，或称为$A$与$B$的和集	
$\bigcup\limits_{i=1}^{n}$	诸并集	unions	$\bigcup\limits_{i=1}^{n}A_i=A_1\cup A_2\cup\cdots\cup A_n$，即诸集$A_1,A_2,\cdots,A_n$的并集	亦可用$\bigcup_{i=1}^{n}$，$\bigcup\limits_{i\in I}$或$\bigcup_{i\in I}$等记法，其中I表示指标集
$\cap$	交集	intersection	$A\cap B=\{x\mid x\in A\wedge x\in B\}$，称为$A$与$B$的交集	
$\bigcap\limits_{i=1}^{n}$	诸交集	intersections	$\bigcap\limits_{i=1}^{n}A_i=A_1\cap A_2\cap\cdots\cap A_n$，即诸集$A_1,A_2,\cdots,A_n$的交集	亦可用$\bigcap_{i=1}^{n}$，$\bigcap_{i\in I}$或$\bigcap\limits_{i\in I}$等记法，其中I为指标集
$\dotplus$	集合的直和	direct sum of sets	若集合A与B不相交，则A与B的并集$A\cup B$称为A与B的直和，记为$A\dotplus B$	亦称不交并
$\dot{\Sigma}$	广义直和	generalized direct sum	若f是标号集A到集族$\{X\}$的一一对应$(f:a\to X_a)$，且当$a\neq b$时，总有$X_a\cap X_b=\varnothing$，则记为$\dot{\sum\limits_{a\in A}}X_a$，并称为集族$\{X\}$的广义直和	
$\setminus$	差集	difference	$A\setminus B$表示所有属于A但不属于B的元的集，称为A与B的差集	
$\triangle$	对称差	symmetric difference	$A\triangle B=(A\setminus B)\cup(B\setminus A)$称为$A,B$的对称差	亦可记为$A\dot{-}B$或$A\ominus B$
U	全集	total set	$A=U$表示A为全集，即全集中所有元素x都属于A	亦可用IΩV表示
$\complement$	余集，补集	complementary set	$\complement_U A=\{x\mid x\in U\wedge x\notin A\}$，即全集$U$中子集$A$的余集或补集	亦可用$\complement A$表示．曾用A^c表示
$\langle\ ,\ \rangle$	有序偶，偶	ordered pair	$\langle a,b\rangle$表示a,b的有序偶	亦可记为(a,b)
$\langle,\cdots,\rangle$	有序元组	elements of ordered	$\langle a_1,a_2,\cdots,a_n\rangle$称为有序$n$元组	亦可记为$(a_1,a_2,\cdots,a_n)$
$\times$	笛卡儿积	Cartesian product	$A\times B=\{(a,b)\mid a\in A,b\in B\}$称为$A$与$B$的笛卡儿积或卡氏积，	$\overbrace{A\times A\times A\times\cdots\times A}^{n个}$记为$A^n$．亦称直积
card	基数，势	cardinal number	card(A)表示集A中诸元的个数，称为A的基数或势	亦可记为$\overline{\overline{A}}$或$\lvert A\rvert$
$\aleph_0$	基数，势	cardinal number	$\aleph_0$表示无限可数集的基数	是希伯来文第一个字母，读Alef
$\sim$	对等	equivalent	$A\sim B$表示集A与集B对等	

符号	中文名称	英文名称	意义或举例	备注
$\mapsto$	元素间的对应	correspond between to elements	在映射下元素间的对应符号，例如，整数集的映射 $\varphi(x)=x^2$ 可表示成 $\varphi:x\mapsto x^2$	
$\to$	映射	mapping	$f:A\to B$ 或 $A\xrightarrow{f}B$ 表示 f 是集 A 到集 B 的映射	
f^{-1}	逆映射	inverse mapping	设 f 是集 A 到 B 的一个双射，则用 f^{-1} 表示 B 到 A 的 f 的逆映射. $f^{-1}f$ 是 A 的恒等映射	亦可用 f_l^{-1}，f_r^{-1} 表示左、右逆映射
R	关系	relation	aRb 表示 a 与 b 有关系 R	
$\not R$	无关系	non-relation	$a\not R b$ 表示 a 与 b 没有关系 R	亦称关系补
$\overline{R}$	反关系	anti-relation	对于二元关系 $R\subseteq X\times Y$，称 $\overline{R}=X\times Y-R$ 为 R 的反关系	亦称否定关系、补关系
R^{-1}	逆关系	inverse relation	对于二元关系 $R\subseteq X\times Y$，称 $R^{-1}\subseteq Y\times X$ 为 R 的逆关系	当且仅当 xRy 时有 $yR^{-1}x$
[ ]	等价类	equivalent class	设 R 是集合 A 上的等价关系，$x\in A$，则称 $[x]_R$ 为 R 的等价类，它是由 A 中那些能使 xRy 成立的所有元素 y 组成的子集	
/	商集	quotient set	设 R 为集 A 的一个等价关系，则商集 A/R 即由一切等价类组成的集合	
$\mathscr{P}$ 或 $\mathfrak{P}$	幂集	power set	用 $\mathscr{P}A$ 或 $\mathfrak{P}A$ 表示集 A 的所有子集组成的集，称为 A 的幂集	
$f\vert_B$	收缩，限制	restriction	设 f 是集合 A 上的一个映射，$B\subseteq A$，则 f 也可看成 B 上的一个映射称为 f 在 B 上的限制或收缩	
$\circ$	合成，复合	composite	$g\circ f$ 表示映射 f 和 g 的合成或复合	
limsup	上极限	superior limit	limsup A_n 表示序列 A_n 的上极限	亦可记为 $\overline{\lim}$
liminf	下极限	inferior limit	liminf A_n 表示序列 A_n 的下极限	亦可记为 $\underline{\lim}$
lim	极限	limit	limA_n 表示序列 A_n 的极限	
$\varinjlim$	归纳极限	inductive limit	$\varinjlim A_\lambda$ 表示 A_λ 的归纳极限	
$\varprojlim$	射影极限	projective limit	$\varprojlim A_\lambda$ 表示 A_λ 的射影极限	
dom	定义域	domain of definition	若 f 为从 A 到 B 的一个映射，则称 A 为映射 f 的定义域，记为 dom f	亦可记为 D(f)
ran f	值域	range	$f(A)=\mathrm{ran}f$. 若 f 为从 A 到 B 的一个映射，则 $f(A)$ 为映射 f 的值域	亦可记为 R(f) 或记为 ran(f)
fld	关系域	domain of a relation	$\mathrm{fld}R=\mathrm{dom}R\cup\mathrm{ran}R$，即关系 R 的域等于 R 的定义域和值域的并集	
codom	陪域	co-domain	若 f 是从集 A 到集 B 的一个映射，则称集 B 是映射 f 的陪域，记为 $B=\mathrm{codom}f$	亦称上域
Imf	像	image	设 f 是集 A 到集 B 的一个映射，用 Imf 表示 A 中所有元素的像构成的集，称为 f 的像集	
$f^{-1}(\ )$	全原像	all inverse image	设 f 是集 A 到集 B 的一个映射，B 中元素 b 的全体逆像组成的集合 $f^{-1}(b)$，称为 b 的全原像	亦称原像
$\preccurlyeq$	弱序关系	weak order relation	$a\preccurlyeq b$, $a,b\in A$ 即集 A 存在弱序关系	
$\prec$	强序关系	strong order relation	$a\prec b$, $a,b\in A$ 即集 A 存在强序关系	
I_A	恒等映射	identity mapping	表示集 A 的每个元素都对应到自身的映射，称为恒等映射	亦称恒等对应. 亦可记为 e_A 或 id A
$\hookrightarrow$; em	嵌入映射	embedding	$A\hookrightarrow B$ 或 emAB 表示 $A\to B$ 的嵌入映射	
n_R	自然映射	natural mapping	n_R 把 A 的一个元素 a 映射成它的等价类 $[a]_R$	亦称正规映射，典则映射
$ub_R(B)$	B 的上界	upper bound of B	$a=ub_R(B)$ 表示 a 是 B 的上界，B 是半序集的子集	
$Lb_R(B)$	B 的下界	lower bound of B	$a=Lb_R(B)$ 表示 a 是 B 的下界，B 是半序集的子集	
ord	一切序数的类	class of every ordinals	表示一切序数构成的类	
cf	共尾度	cofinality	cf α 表示 α 的共尾度	
$K^{<k}$	强极限基数	strong cardinal number of the limit	$K^{<k}=\lim\limits_{\alpha\to K}K^{\alpha}$，其中 K 为正则的强极限基数	

几何与拓扑(Geometry & Topology)

符号	中文名称	英文名称	意义或举例	备注
$\overline{AB}$, AB	[直]线段 AB	segment	表示自点 A 到点 B 的直线段	"直"常略去不写
∠	角	angle	∠AOB 表示角 AOB	
∡	有向角	directed angle	∡ AOB 表示有向角 AOB	
°	度	degree	21°表示 21 度	
′	分	minute	21°13′表示 21 度 13 分	
″	秒	second	21°13′23″表示 21 度 13 分 23 秒	
⌢	弧	arc	$\overset{\frown}{AB}$表示弧 AB. 当$\overset{\frown}{AB}$ 为圆弧时,可用$\overset{\frown}{AB}$° 表示圆弧 AB 对应的度数	
rad	弧度	radian	rad1, radπ 分别表示 1 弧度、π 弧度	rad1≈57°17′45″; radπ=180°
⁻	密位	mil	例如,25⁻, 274⁻表示 25 密位,274 密位	常用在军事数学中度量角的单位符号
π	圆周率	ratio of the circumference of a circle to its diameter	π≈3.141 592 6…表示圆周长与直径的比	英文名称亦可简记为 number π
Rt∠	直角	right angle	等于 90°的角称为直角,记为 Rt∠=90°	曾经记为 rt∠或R∠
△	三角形	triangle	△ABC 表示 A,B,C 三点连线构成的三角形	
⊿	直角三角形	right angle triangle	⊿ ABC 表示直角三角形 ABC	亦可记为 Rt△ABC
▱	平行四边形	parallelogram	▱$ABCD$ 表示平行四边形 $ABCD$	
▭	矩形	rectangle	▭$ABCD$ 表示矩形 $ABCD$	
□	正方形	square	□$ABCD$ 表示正方形 $ABCD$	
⏢	四边形	tetragon	⏢ $ABCD$ 表示任意四边形 $ABCD$	任意二字常略去
◇	菱形	rhombus	◇$ABCD$ 表示菱形 $ABCD$	又名 diamond
⊙	圆	circle	⊙O 表示圆 O	
r,R	半径	radius	从圆心到圆周上任一点的线段称圆的半径,常用 r 或 R 表示	
d,D	直径	diameter	过圆心作任意一条直线,圆内部分的线段称该圆的直径,常用 d 或 D 表示	
C	周长	perimeter	若圆的半径为 r,则周长 $C = 2\pi r$	
//	平行	parallel	AB//CD 表示线段 AB 平行于 CD	
∦	不平行	non-parallel	AB ∦CD 表示直线 AB 与 CD 不平行	
$\underline{//}$	平行且相等	parallel and equal	$AB\underline{//}CD$ 表示线段 AB 与 CD 平行且相等	
⊥	垂直	perpendicular	AB⊥CD 表示线段 AB 垂直于 CD	
≌	全等	congruence	△ABC≌△DEF 表示△ABC 全等于△DEF	
∽	相似	similar	△ABC∽△DEF 表示△ABC 相似于△DEF	
∵	因为	because	∵代表"因为"二字	
∴	所以	therefore	∴代表"所以"二字	
⊻	等角多边形	equiangular polygon	⊻$AB\cdots E$ 表示等角多边形 $AB\cdots E$	多边两字可被省略
⊥̲	等边多边形	equilateral polygon	⊥̲$AB\cdots E$ 表示等边多边形 $AB\cdots E$	多边两字可被省略
α-MN-β	二面角	dihedral angle	平面 α 和平面 β 相交于直线 MN 所成的角	
P-$AB\cdots E$	棱锥	pyramid	顶点是 P、底面多边形是 $AB\cdots E$ 的棱锥	
$AB\cdots E$-$A'B'\cdots E'$	棱柱	prism	上底面是多边形 $AB\cdots E$,下底面是多边形 $A'B'\cdots E'$ 的棱柱	长方体、棱台的记法和此记法类似

符号	中文名称	英文名称	意义或举例	备注
S	面积	area	$S_{\triangle ABC}$ 表示 $\triangle ABC$ 的面积；$S_{球冠}$ 表示某个球冠的面积	
V	体积	volume	$V_{P\text{-}ABC}$ 表示三棱锥 $P\text{-}ABC$ 的体积；$V_{拟柱体}$ 表示某个拟柱体的体积	
$\lvert\ \rvert$	距离	distance	$\lvert AB\rvert$ 表示 A,B 两点间的距离或 AB 线段的长	亦可用 AB 或小写的拉丁字母表示
sin	正弦	sine	$\sin x$ 为 x 的正弦函数	
cos	余弦	cosine	$\cos x$ 为 x 的余弦函数	
tan	正切	tangent	$\tan x$ 为 x 的正切函数	亦可用 $\mathrm{tg}\, x$ 表示
cot	余切	cotangent	$\cot x$ 为 x 的余切函数	亦可用 $\mathrm{ctg}\, x$ 表示
sec	正割	secant	$\sec x$ 为 x 的正割函数	
csc	余割	cosecant	$\csc x$ 为 x 的余割函数	曾用 $\mathrm{cosec}\, x$ 表示
vers	正矢	versedsine	$\mathrm{vers}\, x$ 为 x 的正矢函数	$\mathrm{vers}\, x=1-\cos x$，现已不用
covers	余矢	coversedsine, versedcosine	$\mathrm{covers}\, x$ 为 x 的余矢函数	$\mathrm{covers}\, x=1-\sin x$，现已不用
$\sin^m x$	正弦函数的 m 次方	sine function to the m-th power	$\sin^3 x$ 为 $\sin x$ 的立方	其他三角函数和双曲函数的 m 次方的表示法类似
$\arcsin x$	反正弦主值	principal value of inverse sine	$y=\arcsin x\left(-\frac{\pi}{2}\leqslant y\leqslant\frac{\pi}{2}\right)$	一般值表示成 $\mathrm{Arcsin}x$
$\arccos x$	反余弦主值	principal value of inverse cosine	$y=\arccos x\ (0\leqslant y\leqslant\pi)$	一般值表示成 $\mathrm{Arccos}\, x$
$\arctan x$	反正切主值	principal value of inverse tangent	$y=\arctan x\left(-\frac{\pi}{2}<y<\frac{\pi}{2}\right)$	一般值表示成 $\mathrm{Arctan}x$
$\mathrm{arccot}\, x$	反余切主值	principal value of inverse cotangent	$y=\mathrm{arccot}\, x\ (0<y<\pi)$	一般值表示成 $\mathrm{Arccot}x$
$\mathrm{arcsec}\, x$	反正割主值	principal value of inverse secant	$y=\mathrm{arcsec}\, x\left(0\leqslant y\leqslant\pi，且\ y\neq\frac{\pi}{2}\right)$	一般值表示成 $\mathrm{Arcsec}x$
$\mathrm{arccsc}\, x$	反余割主值	principal value of inverse cosecant	$y=\mathrm{arccsc}\, x\left(-\frac{\pi}{2}\leqslant y\leqslant\frac{\pi}{2}，且\ y\neq 0\right)$	一般值表示成 $\mathrm{Arccsc}x$
T	周期	periodic	$f(x+T)=f(x)$，T 为最小正周期. $T=\pi$ 表示以 π 为周期	
x,y,z	笛卡儿坐标	Cartesian coordinates	$\boldsymbol{e}_x$，$\boldsymbol{e}_y$ 与 $\boldsymbol{e}_z$ 及 $\boldsymbol{r}=x\boldsymbol{e}_x+y\boldsymbol{e}_y+z\boldsymbol{e}_z$ 组成范化正交右手坐标系	
ρ,φ,z	圆柱坐标	cylindrical coordinates	圆柱坐标与笛卡儿坐标的关系为 $x=\rho\cos\varphi, y=\rho\sin\varphi, z=z$	
r,θ,φ	球面坐标	spherical coordinates	球面坐标与笛卡儿坐标的关系为 $x=r\sin\theta\cos\varphi, y=r\sin\theta\sin\varphi, z=r\cos\theta$	
$\boldsymbol{a},\vec{a}$	向量或矢量 $\boldsymbol{a}$	vector $\boldsymbol{a}$	常用 x,y,z 或 x_1,x_2,x_3 表示笛卡儿坐标，则 $\boldsymbol{a}=x\boldsymbol{e}_x+y\boldsymbol{e}_y+z\boldsymbol{e}_z$，简记为 $\boldsymbol{a}=x_i\boldsymbol{e}_i$	印刷常用黑体 $\boldsymbol{a}$，书写常用 $\vec{a}$ 表示
$\lvert\boldsymbol{a}\rvert$	向量的模（绝对值，长度）	module of a vector (absolute value, length)	向量 $\overrightarrow{M_1M_2}$，$\boldsymbol{a}$，$\vec{a}$ 的模依次记为 $\lvert\overrightarrow{M_1M_2}\rvert$，$\lvert\boldsymbol{a}\rvert$，$\lvert\vec{a}\rvert$. 向量的大小称为向量的模	
$\overrightarrow{AB}$	向量 AB	vector AB	表示始点为 A，终点为 B 的向量或有向线段	
$\boldsymbol{e}_a$	单位向量	unit vector	$\boldsymbol{e}_a=a/\lvert a\rvert$ 表示 a 方向的单位向量	亦称幺向量
$\boldsymbol{e}_x$，$\boldsymbol{e}_y$，$\boldsymbol{e}_z$ $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}$	在笛卡儿坐标轴方向的单位向量	unit vector on the Cartesian axial coordinates	$[O;\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}]$表示直角标架；$[O;\boldsymbol{e}_x,\boldsymbol{e}_y,\boldsymbol{e}_z]$表示仿射标架，其中 O 为坐标原点，$\boldsymbol{i},\boldsymbol{j},\boldsymbol{k},\boldsymbol{e}_x,\boldsymbol{e}_y,\boldsymbol{e}_z$ 为基向量	
$\boldsymbol{a}_x$，$\boldsymbol{a}_y$，$\boldsymbol{a}_z$	向量 $\boldsymbol{a}$ 的笛卡儿分量	Cartesian component of a vector $\boldsymbol{a}$	设 $\boldsymbol{a}=\boldsymbol{a}_x+\boldsymbol{a}_y+\boldsymbol{a}_z$，其中 $\boldsymbol{a}_x=x\boldsymbol{e}_x,\boldsymbol{a}_y=y\boldsymbol{e}_y,\boldsymbol{a}_z=z\boldsymbol{e}_z$ 称为向量 $\boldsymbol{a}$ 的笛卡儿分量	
$\boldsymbol{a}\cdot\boldsymbol{b}$ 或 $\boldsymbol{ab}$	标量积或数量积、内积、点积	scalar product, inner product, dot product	$\boldsymbol{a}\cdot\boldsymbol{b}=a_xb_x+a_yb_y+a_zb_z$；$\boldsymbol{a}\cdot\boldsymbol{b}=a_ib_i\overset{\text{def}}{=\!=}\sum a_ib_i$；$\boldsymbol{a}\cdot\boldsymbol{a}=a^2=\lvert\boldsymbol{a}\rvert^2$	亦可表示成 $(\boldsymbol{a},\boldsymbol{b})$，$\langle\boldsymbol{a},\boldsymbol{b}\rangle$，$[\boldsymbol{a},\boldsymbol{b}]$

符 号	中 文 名 称	英 文 名 称	意 义 或 举 例	备 注
$\boldsymbol{a}\times\boldsymbol{b}$	向量积、外积、叉积	vector product, exterior product, cross product	$\boldsymbol{a}\times\boldsymbol{b}$ 是垂直于 $\boldsymbol{a},\boldsymbol{b}$ 所决定平面的向量，且 $\{\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}\times\boldsymbol{b}\}$ 三矢量成右手系. $\lvert\boldsymbol{a}\times\boldsymbol{b}\rvert = \lvert\boldsymbol{a}\rvert\lvert\boldsymbol{b}\rvert\sin(\widehat{\boldsymbol{a},\boldsymbol{b}})$，其中 $(\widehat{\boldsymbol{a},\boldsymbol{b}})$ 表示 $\boldsymbol{a},\boldsymbol{b}$ 的夹角	
$(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ $\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})$	混合积	mixed product	向量 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 的混合积定义为由 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 三向量为邻边组成的平行六面体的有向体积	亦可表示成 $(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$
k	斜率	gradient	直线 $y=kx+b$ 中，k 称为斜率	
e	离心率	eccentricity	在圆锥曲线的极坐标方程中，$r=\dfrac{p}{1-e\cos\varphi}$，$e$ 称为离心率	亦称偏心率.
a	半长轴	semimajor axis	椭圆 $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1(a>b)$ 中，a 称为半长轴	
b	半短轴	semiminor axis	椭圆 $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1(a>b)$ 中，b 称为半短轴	
$\boldsymbol{V}\otimes\boldsymbol{W}$	向量空间的张量积	tensor product of vector spaces	若 $\boldsymbol{V}$ 是 n 维向量空间，$\boldsymbol{W}$ 是 m 维向量空间，则 $\boldsymbol{V}\otimes\boldsymbol{W}$ 是 $n\times m$ 维向量空间的二阶张量	
T_s^r	张量	tensor	设 $\boldsymbol{V}$ 是 n 维向量空间，其对偶空间的二阶张量为 $\boldsymbol{V}^*$，张量积 $V_s^r=\underbrace{\boldsymbol{V}\otimes\cdots\otimes\boldsymbol{V}}_{r个}\otimes\underbrace{\boldsymbol{V}^*\otimes\cdots\otimes\boldsymbol{V}^*}_{s个}$ 的元素称为 (r,s) 型张量	
$\boldsymbol{V}\otimes\boldsymbol{W}$	群的张量积	tensor product of groups	设 $\boldsymbol{V},\boldsymbol{W}$ 是群，$\boldsymbol{V}\otimes\boldsymbol{W}=F(\boldsymbol{V},\boldsymbol{W})/R(\boldsymbol{V},\boldsymbol{W})$ 称为 $\boldsymbol{V},\boldsymbol{W}$ 的张量积	
$T_{xx},T_{xy},\cdots,T_{zz};T_{ij}$	二阶张量 $\boldsymbol{T}$ 的笛卡儿分量	Cartesian component of tensor $\boldsymbol{T}$	$\boldsymbol{T}=T_{xx}\boldsymbol{e}_x\boldsymbol{e}_x+T_{xy}\boldsymbol{e}_x\boldsymbol{e}_y+\cdots,T_{xx}\boldsymbol{e}_x\boldsymbol{e}_x$ 为分张量，	
$\boldsymbol{T}\otimes\boldsymbol{S}$	二阶张量积或并矢积	tensor product dyadic product	两个二阶张量 T 与 S 的张量积 $T\otimes S$ 是具有分量 $T_{ij}S_{kl}$ 的四阶张量	
$\boldsymbol{T}\cdot\boldsymbol{S}$	两个二阶张量的内积	inner product	$\boldsymbol{T}\cdot\boldsymbol{S}$ 表示两个二阶张量 $\boldsymbol{T}$ 与 $\boldsymbol{S}$ 的内积. 它是具有分量 $(\boldsymbol{T}\cdot\boldsymbol{S})_{ik}\overset{\text{def}}{=\!=}\sum\limits_j\boldsymbol{T}_{ij}\boldsymbol{S}_{jk}$ 的二阶张量	
$\boldsymbol{T}\cdot\boldsymbol{a}$	矢量对张量的内积	inner product	$\boldsymbol{T}\cdot\boldsymbol{a}$ 表示二阶张量 $\boldsymbol{T}$ 与矢量 $\boldsymbol{a}$ 的内积. 它是具有分量 $(\boldsymbol{T}\cdot\boldsymbol{a})_i\overset{\text{def}}{=\!=}\sum\limits_j\boldsymbol{T}_{ij}\boldsymbol{a}_j$ 的矢量	
$\boldsymbol{T}:\boldsymbol{S}$	标量积	scalar product	$\boldsymbol{T}:\boldsymbol{S}$ 表示两个二阶张量 $\boldsymbol{T}$ 与 $\boldsymbol{S}$ 的标量积. 它具有标量 $(\boldsymbol{T}:\boldsymbol{S})\overset{\text{def}}{=\!=}\sum\limits_i\sum\limits_j T_{ij}S_{ji}$	
$\overline{\overline{\wedge}}$	透视对应	perspective correspondence	点列 $s(A,B,C,\cdots)$ 与线束 $S(a,b,c,\cdots)$ 是透视的，记为 $s(A,B,C,\cdots)\overline{\overline{\wedge}}S(a,b,c,\cdots)$	
$\overline{\wedge}$	射影对应	projective correspondence	若 $[\pi]$ 与 $[\pi']$ 是两个一维基本形，则它们之间的射影对应记为 $[\pi]\overline{\wedge}[\pi']$	
$\div$	分离	separation	点 A,B 与点 C,D 是分离的，记为 $A,B\div C,D$	
$\overset{-}{\cdot\cdot}$	不分离	nonseparation	点 A,B 与点 C,D 是不分离的，记为 $A,B\ \overset{-}{\cdot\cdot}\ C,D$	
$J,*$	联	join	设 $s=v_0\cdots v_m$ 是 K 的生成复形，$t=w_0\cdots w_n$ 是 L 的生成复形，令 $s*t=v_0\cdots v_m w_0\cdots w_n$，则所有单形 $s*t$ 和它们的面组成的集合是一个单纯复形，称为 K 和 L 的联，记为 $K*L$	亦可记为 $J(K,L)$ 或 KJL
$\boldsymbol{r}=\boldsymbol{r}(t)$	向量函数	vector function	曲线或曲面的参数方程写成向量的形式.	亦称矢函数
$\dfrac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}$ 或 $\boldsymbol{r}'(t)$	导向量	derived vector	$\boldsymbol{r}'(t)=(\boldsymbol{x}'(t),\boldsymbol{y}'(t),\boldsymbol{z}'(t))$ 是向量函数 $\boldsymbol{r}(t)$ 的导向量，有时以弧长 s 为参数的导向量表示成 $\dot{\boldsymbol{r}}(s)$	亦称微商或导矢
$\mathrm{d}\boldsymbol{r}$	微分	differential	设 $\boldsymbol{r}(t)$ 同上，若 $\boldsymbol{r}(t)$ 在 t 处的改变量 $\Delta r=A\Delta t+o(\Delta t)$（$A$ 为固定向量），则称 A 为 $\boldsymbol{r}(t)$ 在 t 点的微分	
$\boldsymbol{r}^{(n)}(t)$	n 阶导向量	n-th derivative	$\boldsymbol{r}^{(n-1)}(t)$ 在 t 点的导向量称为 $\boldsymbol{r}(t)$ 在 t 点的 n 阶导向量	
$\mathrm{d}^n\boldsymbol{r}$	n 阶微分	n-th differential	$\mathrm{d}^{n-1}r$ 在 t 点的微分称为 $r(t)$ 在 t 点的 n 阶微分	
$\dfrac{\partial\boldsymbol{r}}{\partial x_i}$	偏导向量	partial derived vector	若 $\boldsymbol{r}(\boldsymbol{u},\boldsymbol{v})=(\boldsymbol{x}(\boldsymbol{u},\boldsymbol{v}),\boldsymbol{y}(\boldsymbol{u},\boldsymbol{v}),\boldsymbol{z}(\boldsymbol{u},\boldsymbol{v}))$，则 $\boldsymbol{r}_u(\boldsymbol{u},\boldsymbol{v})=\left(\dfrac{\partial\boldsymbol{x}}{\partial\boldsymbol{u}},\dfrac{\partial\boldsymbol{y}}{\partial\boldsymbol{u}},\dfrac{\partial\boldsymbol{z}}{\partial\boldsymbol{u}}\right)$ 是 $\boldsymbol{r}(\boldsymbol{u},\boldsymbol{v})$ 关于 $\boldsymbol{u}$ 的偏导向量	亦称偏导矢

符号	中文名称	英文名称	意义或举例	备注
$\boldsymbol{T}(s)$	单位切向量	unit tangent vector	$\boldsymbol{T}(s)=\dot{\boldsymbol{r}}(s)$表示曲线 C 在一点处的单位切向量，其中 s 为曲线 C 的弧长参数	亦可表示成$\boldsymbol{\alpha}(s)$
$\boldsymbol{N}(s)$	主法向量	principal normal vector	$\boldsymbol{N}(s)=\dfrac{\ddot{\boldsymbol{r}}(s)}{\|\ddot{\boldsymbol{r}}(s)\|}$表示曲线 C 在一点处的主法向量，$\boldsymbol{N}(s)$指向曲线 C 凹入的方向	亦可表示成$\boldsymbol{\beta}(s)$
$\boldsymbol{B}(s)$	副法向量	binormal vector	$\boldsymbol{B}(s)=\boldsymbol{T}(s)\times\boldsymbol{N}(s)$表示曲线 C 在一点处的副法向量	亦称从法向量.表示成 $\gamma(s)$
$\{P;\boldsymbol{T};\boldsymbol{N},\boldsymbol{B}\}$	活动标架	Frenet frame	$\boldsymbol{T},\boldsymbol{N},\boldsymbol{B}$ 依次构成右手系，它们构成一个标架，称为曲线 C 在 P 点的活动标架或弗雷内标架	
k	曲率	curvature	曲率 k 是表示曲线弯曲程度的量.曲率 k 越大，曲线弯曲程度越大，曲率小，曲线弯曲程度小	直线的曲率为 0
τ	挠率	torsion	挠率是表示空间曲线扭翘程度的量.挠率的绝对值大，曲线扭翘程度大，挠率的绝对值小，曲线扭翘程度小.平面曲线的挠率为 0	
$k_r(s)$	相对曲率	relative curvature	表示平面曲线弯曲程度和弯曲方向的量	
i_r	旋转指标	rotation index	$i_r=\dfrac{1}{2\pi}\int_0^l k_r(s)\mathrm{d}s$ 表示平面闭曲线 C 的旋转指标，它是曲线 C 的切线像($r=\boldsymbol{T}(s)$)在单位圆周上环绕的圈数	若 C 是平面简单闭曲线，则 $i_r=\pm 1$
$\boldsymbol{n}$	单位法向量	unit normal vector	曲面 $\boldsymbol{r}=\boldsymbol{r}(a,v)$ 上一点 $P(u,v)$ 处的单位法向量 $\boldsymbol{n}=\dfrac{\boldsymbol{r}_u\times\boldsymbol{r}_v}{\|\boldsymbol{r}_u\times\boldsymbol{r}_v\|}$	式中各量均在(u,v)取值.$\boldsymbol{r}_u,\boldsymbol{r}_v,\boldsymbol{n}$ 依序构成右手系
E,F,G,g_{ij}	曲面的第一类基本量	fundamental quantities of first kind for surfaces	对曲面 $\boldsymbol{r}=\boldsymbol{r}(u,v)$，其第一类基本量分别为 $E=\boldsymbol{r}_u\cdot\boldsymbol{r}_u,F=\boldsymbol{r}_u\cdot\boldsymbol{r}_v,G=\boldsymbol{r}_v\cdot\boldsymbol{r}_v.$ $g_{ij}=\boldsymbol{r}_i\cdot\boldsymbol{r}_j\quad(i,j=1,2)$	$E>0,\ G>0,$ $EG-F^2>0$
Ⅰ	曲面的第一基本形式	first fundamental form of a surface	$\mathrm{I}=E\mathrm{d}u^2+2F\mathrm{d}u\mathrm{d}v+G\mathrm{d}v^2$	第一基本形式是正定的，它决定曲面的内蕴性质
L,M,N,L_{ij}	曲面的第二类基本量	fundamental quantities of second kind for surfaces	对曲面 $\boldsymbol{r}=\boldsymbol{r}(u,v)$，其第二类基本量分别为 $L=\boldsymbol{r}_{uu}\cdot\boldsymbol{n},M=\boldsymbol{r}_{uv}\cdot\boldsymbol{n},N=\boldsymbol{r}_{uv}\cdot\boldsymbol{n}.$ $L_{ij}=\boldsymbol{r}_{ij}\cdot\boldsymbol{n}\ (i,j=1,2)$	
Ⅱ	曲面的第二基本形式	second fundamental form of a surface	$\mathrm{II}=L\mathrm{d}u^2+2M\mathrm{d}u\mathrm{d}v+N\mathrm{d}v^2$	
k_n	法曲率	normal curvature	曲面 S 在 P 点沿方向 $\boldsymbol{a}$ 的法截线曲率可作为曲面在该点的法曲率 k_n	其绝对值相等
K_c	全曲率	total curvature	$K_c=\int_0^l k(s)\mathrm{d}s$ 表示曲线 C 的全曲率	
K_r	相对全曲率	relative total curvature	$K_r=\int_0^l k_r(s)\mathrm{d}s$ 表示曲线 C 的相对全曲率	
K	总曲率	Gaussian curvature	$K=k_1k_2$ 表示曲面 S 在点 P 的弯曲情况.曲面上的点可按总曲率的符号进行分类.$K>0$ 的点是椭圆点，$K<0$ 的点是双曲点，$K=0$ 的点是抛物点	亦称高斯曲率.式中 k_1,k_2 为其对应的主曲率
H	平均曲率	mean curvature	表示曲面 S 在点 P 的平均曲率	亦称中曲率
e,f,g	曲面的第三类基本量	fundamental quantities of third kind for surfaces	对曲面 $\boldsymbol{r}=\boldsymbol{r}(u,v)$，其第三类基本量分别为 $e=\boldsymbol{n}_u\cdot\boldsymbol{n}_u,f=\boldsymbol{n}_u\cdot\boldsymbol{n}_v,g=\boldsymbol{n}_v\cdot\boldsymbol{n}_v$	
Ⅲ	曲面的第三基本形式	third fundamental form of a surface	$\mathrm{III}=\mathrm{d}\boldsymbol{n}\cdot\mathrm{d}\boldsymbol{n}=e\mathrm{d}u^2+2f\mathrm{d}u\mathrm{d}v+g\mathrm{d}v^2$	
$[jk,i]$	第一类克里斯托费尔符号	Christoffel symbol of the 1st kind	$[jk,i]=\dfrac{1}{2}\left(\dfrac{\partial g_{ij}}{\partial x^k}+\dfrac{\partial g_{ki}}{\partial x^j}-\dfrac{\partial g_{jk}}{\partial x^i}\right)$	亦可表示成 Γ_{jki}
$\left\{\begin{matrix}k\\ij\end{matrix}\right\}$	第二类克里斯托费尔符号	Christoffel symbol of the 2nd kind	$\left\{\begin{matrix}k\\ij\end{matrix}\right\}=\dfrac{1}{2}g^{kl}\left(\dfrac{\partial g_{lj}}{\partial x^i}+\dfrac{\partial g_{il}}{\partial x^j}-\dfrac{\partial g_{ij}}{\partial x^l}\right)$	亦可表示成 $\Gamma_{ij}^k=g^{kl}\Gamma_{ijl}\Gamma_{ij}^k$，也称为联络系数
k_g	测地曲率	geodesic curvature	曲面 S 上的曲线 C 在某一点 P 的切平面上的投影线的曲率可作为曲线 C 的测地曲率	其绝对值相等
exp	指数映射	exponential map	指数映射 $\exp:T_P\to S$ 是曲面 S 上 P 的切平面 T_P 的切向量与曲面 S 上点的对应关系.若 $\boldsymbol{v}\in T_P$，过 P 沿 $\boldsymbol{v}$ 的方向作测地线 C，在 C 上取点 M，使 $\widehat{PM}=\|\boldsymbol{v}\|$，则 $\exp\boldsymbol{v}=M$	

符号	中文名称	英文名称	意义或举例	备注
τ_g	测地挠率	geodesic torsion	在曲面S上过一点P作以单位切向量α为初始方向的测地线$C:u=u(s),v=v(s)$，测地线C在P点的挠率称为曲面S在P点关于α方向的测地挠率	$\tau_g=\left(\alpha,n,\frac{\mathrm{d}n}{\mathrm{d}s}\right)$
$\mathscr{N}$	高斯映射	Gauss map	以曲面S的单位法向量$\boldsymbol{n}(u,v)$作为向量函数，表示单位球面S^2，高斯映射$\mathscr{N}:S\to S^2$是曲面S与相应的球面S^2之间的对应关系	亦称曲面的球面表示
$\deg\mathscr{N}$	高斯映射度	Gauss mapping degree	$\deg\mathscr{N}=\frac{1}{2}\chi(S)$表示高斯映射度，它由曲面拓扑所决定，其中$\chi(S)$表示欧拉示性数	
$\{U_\alpha,\varphi_\alpha\}$	坐标邻域	coordinate neighborhoods	U_α是微分流形M的开集，φ_α是微分流形U_α到R^n的开子集的同胚	
C^∞	C^∞相容	C^∞ compatible	$U\cap V\neq\varnothing$，$\varphi\circ\psi^{-1}$和$\psi\circ\varphi^{-1}$是R^n的开子集$\varphi(U\cap V)$和$\psi(U\cap V)$的C^∞微分同胚. 称(U,φ)和(V,ψ)是C^∞相容的	
L_XY	李导数	Lie derivative	$(L_XY)_p=\lim\limits_{t\to 0}\frac{1}{t}((\varphi_{-t})_*Y_{\varphi_t(p)}-Y_p)$ $=\frac{\mathrm{d}}{\mathrm{d}t}(\varphi_{-t})_*Y_{\varphi_t(p)}\mid_{t=0}$	
R^l_{ijk}	黎曼曲率张量	Riemannian curvature tensor	$R^l_{ijk}=\frac{\partial}{\partial x^i}\Gamma^l_{jk}-\frac{\partial}{\partial x^j}\Gamma^l_{ik}+\Gamma^l_{ih}\Gamma^h_{jk}-\Gamma^l_{jh}\Gamma^h_{ik}$和$R_{ijkl}=R^h_{ikl}$ g_K均称为黎曼曲率张量	亦称第二类克里斯托费尔符号
Ric	里奇曲率张量	Ricci curvature tensor	$\mathrm{Ric}(X,Y)=\sum\limits_i R(e_i,X,Y,e_i)$，即里奇曲率张量是一个(0,2)型张量场. 由对称性知 $\mathrm{Ric}(X,Y)=\mathrm{Ric}(Y,X)$	
C_{ijkl}	共形曲率张量	conformal curvature tensor	$C_{ijkl}=R_{ijkl}-\frac{1}{n-2}\{R_{ik}g_{il}-R_{il}g_{il}+R_{jl}g_{ik}-R_{jk}g_{il}\}+\frac{s}{(n-1)(n-2)}(g_{ik}g_{il}-g_{il}g_{jk})$	亦称外尔张量
P^l_{ijk}	射影曲率张量	projective curvature tensor	$P^l_{ijk}=R^l_{ijk}-\frac{1}{n-1}(\delta^l_kR_{ij}-\delta^l_jR_{ik})$称为射影曲率张量	
d	外微分算子	exterior differential operator	对于任意$\omega_1,\omega_2\in A^p(M)$：1. $\mathrm{d}(\omega_1+\omega_2)=\mathrm{d}\omega_1+\mathrm{d}\omega_2$；2. $\mathrm{d}(\omega_1\wedge\omega_2)=\mathrm{d}\omega_1\wedge\omega_2+(-1)^p\omega_1\wedge\mathrm{d}\omega_2$；3. 若$f\in A^0(M)$，则$\mathrm{d}(\mathrm{d}f)=0$	若$f\in A^0(M)$，则$\mathrm{d}f$恰是f的微分
$Z^p(M,R)$	光滑p次闭形式空间	space of smooth p-closed differential form	$Z^p(M,R)=\{\omega\mid\omega$是流形$M$上的光滑$p$次闭形式$\}$表示光滑$p$次闭形式空间	
$B^p(M,R)$	光滑p次恰当形式空间	space of smooth p-exact differential form	$B^p(M,R)=\{\omega\mid\omega$是流形$M$上的光滑$p$次恰当形式$\}$表示光滑$p$次恰当形式空间	
$H^p(M,R)$	德·拉姆上同调群	de Rham cohomology group	表示流形M的第p个德·拉姆上同调群. $H^p(M,R)$中的元素称为同调类	亦称第p个德·拉姆上同调空间
$\int_M\omega$	形式积分	integral of forms	$\int_M\omega=\sum\limits_i\int_M f_i\circ\omega$	
∇	仿射联络	affine connection	设M是n维C^∞流形，$\Gamma(TM)$为M上的C^∞向量场空间. M上的仿射联络是指映射$\nabla:\Gamma(TM)\times\Gamma(TM)\to\Gamma(TM)$，满足四条公理	
$\nabla_{X_p}Y$	共变导数	covariant derivative	令$P\in M,X_P\in T_P(M)$. Y为M上的C^∞向量场. 定义$\nabla_{X_P}Y=(\nabla_XY)_P$	亦称协变微商
$K(X,Y)$	截面曲率	sectional curvature	对任意两个不共线的切向量$X,Y\in T_PM$， $K(X,Y)=-\frac{R(X,Y,X,Y)}{g(X,X)g(Y,Y)-[g(X,Y)]^2}$	当$\dim M=2$时，$K(X,Y)$恰好是M在P点的高斯曲率
$R(X,Y)$	曲率算子	curvature operator	$R(X,Y)Z=\nabla_X(\nabla_YZ)-\nabla_Y(\nabla_XZ)$ $-\nabla_{[X,Y]}Z\ (X,Y,Z\in\Gamma(TM))$	
$\triangle$	拉普拉斯-贝尔脱拉米算子	Laplace-Bertrami operator	$\triangle f=\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^i}\left(\sqrt{g}\,g^{ij}\frac{\partial f}{\partial x^j}\right)$	
$S_p(2n)$	辛群	symplectic group	设(V,ω)是一个辛空间，(V,ω)的自同构的全体构成群$\mathrm{GL}(V)$的一个子群记为$\mathrm{SP}(V,\omega)$，特别地，标准辛空间(K^{2n},ω)的自同构群记为$S_p(2n,K)$. 若$K=R$，则把$S_p(2n,K)$简记为$S_p(2n)$，并称为$2n$维辛群	

符 号	中文名称	英文名称	意 义 或 举 例	备 注
$E(f)$	能量	energy	设 M,N 为黎曼流形，$f:M\to N$ 为光滑映射，f 的能量定义为：$E(f)=\frac{1}{2}\int_M \|df\|^2 *1$，其中 $*1$ 为 M 的体积元	
$e(f)$	能量密度	energy density	符号条件同上，$e(f)=\frac{1}{2}\|df\|^2$	
∂M	流形的边界	boundary of a manifold	带边流形 M 中全体边界点的集	
T_PM	切空间	tangent space	微分流形 M 在 P 点处的全体切向量的集记为 T_PM，称为 M 在 P 处的切空间	T_PM 是实 $\dim M$ 维向量空间
f_{*P},T_Pf	在一点处的切映射	tangent map at a point	$f:M\to N$ 是可微映射，$f_{*P}:T_PM\to T_{f(P)}N$ 称为可微映射 f 在 $P\in M$ 处的切映射	若 f 是微分同胚，则 $\forall P\in M, f_{*P}$ 是同构
TM	流形的切丛	tangent bundle of manifold	(TM,π,M) 称为微分流形 M 的切丛，简称 TM 为 M 的切丛	
Tf	切映射	tangent map	设 $f:M\to N$ 是流形 M 到 N 的可微映射，$Tf:TM\to TN$ 称为 f 的切映射	若 $f:M\to N$ 是微分同胚，则 $Tf:TM\to TN$ 亦然
$\xi\oplus\eta$	向量丛的惠特尼和	Whitney sum of vector bundles	ξ,η 分别是 n 维，k 维向量丛，$\tilde{\pi}:E(\xi)\oplus E(\eta)\to B$ 为自然投射.$(E(\xi)\oplus E(\eta),\tilde{\pi},B)$ 是 $n+k$ 维向量丛，称为 ξ 与 η 的惠特尼和	亦可看成积丛 $\xi\times\eta$ 由对角映射 $f:B\to B\times B$ 决定的诱导丛
$\chi(\xi)$	欧拉数	Euler number	设 $\xi=(E,\pi,M)$ 是 n 维定向向量丛，则零截面的自交数称为向量丛 ξ 的欧拉数	当 $\xi=TM$ 时，$\chi(\xi)$ 就是 M 的欧拉示性数
$U^{\perp}(t)$	正交分量	orthogonal component	表示分向量场 $U(t)$ 与测地线 γ 正交的分量	
$T_x^{\perp}M$	法空间	normal space	表示 M 在 x 处的法空间，正交于切空间 T_xM	
$\nabla^{\perp}$	法联络	normal connection	若 M 是黎曼流形，则 $\nabla^{\perp}$ 表示 M 上的法联络	
$(\tilde{R}(X,Y)Z)^{\perp}$	正交投影	orthogonal projection	表示 $\tilde{R}(X,Y)Z$ 在 M 的法丛 $N(M)$ 上的投影.式中 $\tilde{R}$ 是 $\tilde{M}$ 的曲率张量	
(X,d)	度量空间	metric space	赋予度量 d 的集合 X 称为度量空间	亦称距离空间
$(X,\mathscr{T})$	拓扑空间	topological space	确定了拓扑 $\mathscr{T}$ 的集合 X 称为拓扑空间	
$\bar{A}$，$\mathrm{cl}A$	闭包	closure	包含 A 的所有闭集的交集称为 A 的闭包，它是包含 A 的最小闭集	
$b(A)$，$\mathrm{Bd}A$	边界	boundary	A 的全体边界点组成的集合称为 A 的边界	亦可记为 $A^b,\partial A$
$\mathrm{Int}\ A,A^i$	内部	interior	集 A 的全部内点组成的集合称为 A 的内部	亦可记为 $\mathring{A}$ 或 A°
$U(a,\delta)$	邻域	neighborhood	$U(a,\delta)=\{x\|a-\delta<x<a+\delta\}$ 称为点 a 的 δ 邻域点 a 称为邻域的中心，δ 称为邻域的半径	
$\mathring{U}(a,\delta)$	去心邻域	deleted neighborhood	$\mathring{U}(a,\delta)=\{x\|0<\|x-a\|<\delta\}$ 称为点 a 的去心的 δ 邻域	
$\mathscr{U}(x)$	邻域系	neighborhood system	点 x 的邻域的全体称为 x 的邻域系	
$X\vee Y$	拓扑空间的楔和	wedge sum of topological spaces	设 X,Y 为两个带有基点的拓扑空间.x_0,y_0 分别为 X,Y 的基点.子空间 $X\times\{y_0\}\cup\{x_0\}\times Y\subset X\times Y$ 称为 X 和 Y 的楔和	
$X\wedge Y$	拓扑空间的碎积	smash product of topological spaces	商空间 $X\times Y/X\vee Y$ 称为 X,Y 的碎积	
$V_{n,k}$	斯蒂弗尔流形	Stiefel manifold	$V_{n,k}=\{(e_1,e_2,\cdots,e_k)\|e_i\in R^n,e_i\cdot e_j=\delta_{ij},1\leqslant i,j\leqslant k\}$ 在 $R^n\times\cdots\times R^n$（$k$ 个）的诱导拓扑之下，$V_{n,k}$ 为一个紧致流形，称为斯蒂弗尔流形	
$B_{\varepsilon}(a)$	开球	open ball	设 (X,d) 为度量空间，$a\in X,\varepsilon>0,B_{\varepsilon}(a)=\{x\in X\|d(a,x)<\varepsilon\}$ 称为以 a 为中心的 ε 开球	亦可记为 $B(a,\varepsilon)$
$\bar{B}_{\varepsilon}(a)$	闭球	closed ball	设 (X,d) 为度量空间，$a\in X,\varepsilon>0,\bar{B}_{\varepsilon}(a)=\{x\in X\|d(a,x)\leqslant\varepsilon\}$ 称为以 a 为中心的 ε 闭球	亦可记为 $\bar{B}(a,\varepsilon)$
$\delta(M)$	直径	diameter	设 M 为度量空间 (X,d) 的子集，定义 $\delta(M)=\sup\{d(x,y)\|x,y\in M\}$，称为集 M 的直径	亦可记为 $\mathrm{diam}M$
$A^d,d(A)$	导集	derived set	集 A 的一切聚点的集称为 A 的导集	
$A^e,\mathrm{ext}(A)$	外部	exterior	集 A 的全体外点组成的集合称为 A 的外部，记为 A^e 或 $\mathrm{ext}(A)$	

符号	中文名称	英文名称	意义或举例	备注
$\mathrm{Ind}X$	大归纳维数	large inductive dimension	这是在正则空间中利用归纳法定义的维数，若空间 X,Y 同胚，则 $\mathrm{Ind}X=\mathrm{Ind}Y$	亦称布劳威尔－切赫维数
$\mathrm{ind}\ X$	小归纳维数	small inductive dimension	这是在正则空间中利用归纳法定义的维数，若空间 X,Y 同胚，则 $\mathrm{ind}X=\mathrm{ind}Y$	亦称门杰－乌雷松维数
$\varprojlim\{X_\alpha,\pi_\alpha^\beta,A\}$	逆极限	inverse limit	逆系 $\{X_\alpha,\pi_\alpha^\beta,A\}$ 的逆极限	亦可记为 $\varprojlim X_\alpha$
$\varepsilon(A)$	凸包络	convex envelope	X 内所有包含 A 的凸集之交称为 A 的凸包络	
$\simeq$	同伦	homotopy	若 $f,g:X\to Y$ 都是连续映射，$I=[0,1]$，且存在连续映射 $H:X\times I\to Y$，使得对所有 $x\in X$，$H(x,0)=f(x)$，$H(x,1)=g(x)$，则 f,g 称为同伦映射，记为 $f\simeq g:X\to Y$	这里 H 称为从 f 到 g 的一个同伦或伦移
$\approx$	同胚	homeomorphism	$f:X\to Y$ 是连续映射，且 f 的逆映射连续，则称 f 为同胚，亦称空间 X 与 Y 同胚，记为 $X\approx Y$	亦称拓扑映射、拓扑变换
$\Vert\ \Vert$	范数	norm	$\Vert x\Vert$ 表示赋范空间中 x 的范数或实空间中向量 α 的赋值．记为 $\Vert\alpha\Vert$	欧氏空间的向量 x 的长度概念的推广
E^n	n 维欧氏空间	n-dimensional Euclidean space	$E^n=\{(x_1,x_2,\cdots,x_n)\mid x_i\in\mathrm{R}\}$，规定度量 $d=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$	亦可记为 R^n
P^n	n 维射影空间	n-dimensional projective space	域 F 上的 n 维射影空间常记为 FP^n，简记为 P^n，当 F 是实数域时记为 $\mathrm{R}P^n$；当 F 是复数域时记为 $\mathrm{C}P^n$．若 F 是四元数域 H．记为 $\mathrm{H}P^n$	
S^n	n 维球面	n-dimensional sphere	$S^n=\{x\in\mathrm{R}^{n+1}:\lvert x\rvert=r\}$	
T^n	n 维环面	n-dimensional torus	圆 S^1 自身的 n 次拓扑乘积．记为 $T^n=S^1\times S^1\times\cdots\times S^1$	
$C_q(\ ,\ )$	链群	chain group	K 是复形，$C_q(K,Z)$ 称为 K 的 q 维链群	亦可简记为 $C_q(K)$
H_n	n 维同调群	n-dimensional homology group	$H_n(K,A)=Z_n(K,A)/B_n(K,A)$ 表示复形 K 的以 A 为系数群的 n 维同调群	
H^n	n 维上同调群	n-dimensional cohomology group	$H^n(X,A)=Z^n(K,A)/B^n(K,A)$ 表示复形 K 以 A 为系数群的 n 维上同调群	
$\check{H}^n$	n 维切赫上同调群	n-dimensional Čech cohomology group	$\check{H}^n(X)=\varinjlim H^n(N_\lambda)$ 表示 X 的 n 维切赫上同调群	
$\check{H}_n$	n 维切赫同调群	n-dimensional Čech homology group	$\check{H}_n(X)=\varprojlim H_n(N_\lambda)$ 表示 X 的 n 维切赫同调群	
π_n	n 维同伦群	n-dimensional homotopy group	$\pi_n(X)$ 是映射 $(S^n,s_0)\to(X,x_0)$ 的同伦类集合	
$\pi_{n+k}(S^n)$	稳定同伦群	stable homotopy group	悬垂同态 $E:\pi_{n+k}(S^n)\to\pi_{n+k+1}(S^{n+1})$，当 $n>k+1$ 时为同构，称为球面的第 k 个稳定同伦群	悬垂同态亦称同纬像同态
∂	边缘算子	boundary operator	∂c 表示 c 的边缘	
δ	上边缘算子	coboundary operator	δf 表示 f 的上边缘	
sq	斯廷罗德方形运算	Steenrod square	$Sq^i(x,y)=\sum_{j+k=i}Sq^j(x)Sq^k(y)$ 即 x 的斯廷罗德方形运算	
$\mathscr{P}$	斯廷罗德幂运算	Steenrod power	$\mathscr{P}_p^r(xy)=\sum_{i+j=r}\mathscr{P}_p^i(x)\mathscr{P}_p^j(y)$ 即 x 的斯廷罗德 p 次幂运算	亦可记为 St_p^r
$\smile$	上积	cup product	$z_1\smile z_2$ 表示 z_1 和 z_2 的上积	
$\frown$	卡积	cap product	$z_1\frown z_2$ 表示 z_1 和 z_2 的卡积	
$\omega\wedge\eta$	外积	exterior product	表示微分形式 ω,η 的外积．$\omega\wedge\eta=A_{k+l}(\omega\otimes\eta)$．其中 A_{k+1} 是反对称化算子，ω 是 k 次矢量，η 是 l 次矢量，$\omega\wedge\eta$ 是 $(k+l)$ 次外矢量	
mesh	复形的网径	mesh diameter of a complex	单纯复形 K 中诸单形直径的最大值称为复形的网径，即 $\mathrm{mesh}=\max_{\sigma\in K}\{\Vert x-y\Vert \mid x,\ y\in\sigma\}$	
deg	映射度	degree of mapping	设 $f:S^n\to S^n$ 是映射，α 是 $H_n(S^n)$ 的生成元，则 $f_*(\alpha)=\rho\alpha$．其中整数 ρ 称为 f 的映射度．记为 $\rho=\deg(f)$	亦称拓扑度，又称布劳威尔度
rel	相对于	relative	rel A 表示相对于 A	

符号	中文名称	英文名称	意义或举例	备注
C	连续函数空间	continuous function space	$C[a,b]$表示$[a,b]$上连续函数的全体	
L^p	p次可积函数空间	integrable function space of order p	$L^p(\Omega,\mathscr{B},\mu)(\infty>p\geqslant 1)$是测度空间$(\Omega,\mathscr{B},\mu)$上可测而且$p$次可积函数的全体	
C^n	C^n类函数空间	C^n class function space	$C^n[a,b](\infty>n\geqslant 1)$是$[a,b]$上$n$阶连续可微函数的全体	
C^∞	C^∞类函数	function of class C^∞	对于所有r，函数f是C^r类的．亦称f是光滑的	
C^∞	C^∞映射	C^∞ mapping	W,N是微分流形，$F:W\to N,\psi\circ F\circ\varphi^{-1}P:\varphi(U)\to\psi(V)$是$C^\infty$的．$U,V$分别是$W,N$的坐标邻域	
L^∞	本性有界可测函数	essentially bounded function space	$L^\infty(\Omega,\mathscr{B},\mu)$表示$\Omega$上(关于$\mu$)本性有界可测函数全体	
T_2	豪斯多夫空间	Hausdorff space	设X为拓扑空间，若X的任意两个不相同的点都有不相交的开邻域则称X为豪斯多夫空间	亦称T_2空间
R^∞	希尔伯特空间	Hilbert space	设$x=(x_1,x_2,\cdots),y=(y_1,y_2,\cdots).x,y\in R^\infty$，定义$d=\sqrt{\sum_{i=1}^{\infty}(x_i-y_i)^2}$，则$(R^\infty,d)$称为希尔伯特空间	
Y^X	函数空间	functional space	表示所有连续函数$f:X\to Y$的集合	
$N_{K,U}$	紧致开拓扑	compact open topology	$N_{K,U}=\{f:f(K)\subset U\}$，其中$K\subset X$紧致，$U\subset Y$为开集	
e_α^n	n维胞腔	cell of dimension n	e_α^n是空间X的子集	
CW	CW复形	CW-complex	一个空间X中的CW复形是满足闭包有限和诱导弱拓扑两项条件的胞腔复形	
$L(p,q)$	透镜空间	lens spaces	$L(p,q)=s^3/Zp$	
WHE	弱同伦等价公理	weak homotopy equivalence axiom	若$f:X\to Y$是弱同伦等价关系，则$f_*:k_n(X,x_0)\to k_n(Y,f(x_0))$是同构	
$\widetilde{KO}(X)$	$\widetilde{KO}$群	$\widetilde{KO}$-group	表示X上实向量丛的所有稳定等价类集合	
$\widetilde{K}(X)$	$\widetilde{K}$群	$\widetilde{K}$-group	表示X上复向量丛的所有稳定等价类集合	
$\widetilde{KS}_p(X)$	$\widetilde{KS}_p$群	$\widetilde{KS}_p$-group	表示X上四元向量丛的所有稳定等价类集合	
$K(s)$	K群	K-group	表示由半群的同态$\varnothing:S\to K(s)$诱导的abelian群	
$KO(X)$	KO群	KO-group	$KO(X)\cong\widetilde{KO}(X)\oplus KO(\{x_0\})$	
$K(X)$	K群	K-group	$K(X)\cong\widetilde{K}(X)\oplus K(\{x_0\})$	
$KS_p(X)$	KS_p群	KS_p-group	$KS_p(X)\cong\widetilde{KS}_p(X)\oplus KS_p(\{x_0\})$	
$M_1\sim M_2$	流形的协边	cobordism of manifolds	设M_1,M_2都是紧致(无边)微分流形，若存在紧致带边流形W与微分同胚$\partial W\cong M_1\times(0)\cup M_2\times(1)$，则称$M_1$与$M_2$协边	
MSO_n	定向协边群	oriented bordism group	表示所有定向协边类的集合	亦称Thom群
MO_n	非定向协边群	unoriented bordism group	表示所有非定向协边类的集合	亦称Thom群
MSO_*	分次交换环	graded commutative ring	$MSO_*=\sum MSO_n$	
MO_*	分次交换代数	graded commutative algebra	$MO_*=\sum MO_n$	
$MSO_n(X,A)$	定向奇异协边群	oriented singular bordism group	表示(X,A)中定向奇异协边类的集合	
$MSO_*(X,A)$	分次右模	graded right module	$MSO_*(X,A)=\sum MSO_n(X,A)$	
$MSO_n(Pt)$	一点的协边群	bordism group of a point	$MSO_n(Pt)=MSO_n$	
$\widetilde{MSO}_n(X)$	约化群	reduced group	表示增广同态$\varepsilon_*:MSO_n(X)\to MSO_n(pt)$的核	
$MO_n(X,A)$	非定向奇异协边群	unoriented bordism group	表示(X,A)中非定向奇异协边类的集合	
$MO_*(X,A)$	分次模	graded module	$MO_*(X,A)=\sum MO_n(X,A)$	

代数学(Algebra)

符号	中文名称	英文名称	意义或举例	备注
max	最大或极大	maximum	$y_{max}=a$ 表示 y 的最大(极大)值等于 a	
min	最小或极小	minimum	$y_{min}=b$ 表示 y 的最小(极小)值等于 b	
!	阶乘	factorial	$n!=1\cdot2\cdot3\cdot\cdots\cdot n$	规定 $0!=1$
!!	双阶乘	double factorial	$(2n)!!=2\cdot4\cdot6\cdot\cdots\cdot(2n)$; $(2n+1)!!=1\cdot3\cdot5\cdot\cdots\cdot(2n+1)$	
$(a)_n$	始于 a 的 n 个实数之积	product of the n-real numbers by the beginning at a	例如, $(\sqrt{2})_4=\sqrt{2}(\sqrt{2}+1)(\sqrt{2}+2)(\sqrt{2}+3)$	a 为实数,n 为自然数
C_n^p 或 $\binom{n}{p}$	二项式系数,组合数	binomial coefficient, combinatorial numbers	表示从 n 个元素中每次取出 p 个元素的所有不同组合的总数	
P_m^n 或 A_m^n	选排列	selections permutation	$P_m^n=\dfrac{m!}{(m-n)!}=m(m-1)\cdots(m-n+1)$	
P_m 或 A_m	全排列	all permutation	$P_m=m!$	
H_m^n	重复组合	combination with repetition	$H_m^n=C_{m+n-1}^n=\dfrac{(m+n-1)!}{n!(m-1)!}$	
$\cup_m^n$	有重复的排列	permutation with repetition	$\cup_m^n=m^n$,即从 m 个相异元素中每次取出 n 个元素允许重复排列的排列总数	亦可记为 $\|\|_m^n=m^n$
R_m^n	环排列	circular permutation	$R_m^n=\dfrac{P_m^n}{n}=C_m^n(n-1)!\ (m\leqslant n)$. 当 $m=n$ 时,$R_m^m=(m-1)!$	亦可用 $R_{m平}^n$ 和 $R_{m立}^n$ 分别表示平面环排列与空间环排列
i	虚数单位	imaginary unit	$i=\sqrt{-1}\ (i^2=-1)$	电工技术中常用 j
z	复数记号	symbol of complex number	$z=a+bi$ 即实部为 a,虚部为 b 的复数	
Re z	z 的实部	real part of z	$z=a+bi$ (Re $z=a$)	
Im z	z 的虚部	imaginary part of z	$z=a+bi$ (Im $z=b$)	
$\|z\|$	z 的模	modulus of z	$z=a+bi\ (\|z\|=\sqrt{a^2+b^2})$	亦可用 mod z 表示
arg z	z 的辐角	argument of z	$\varphi=\arg z$ 即复数 z 的辐角为 φ,$0<\varphi\leqslant2\pi$	
$\bar{z}$	z 的共轭复数	conjugate complex number of z	设 $z=a+bi$,则 $\bar{z}=a-bi$ 称为 z 的共轭复数	亦可用 z^* 表示
sgn z	z 的单位模函数	signum z	$\operatorname{sgn} z=\begin{cases}z/\|z\| & (z\neq0),\\0 & (z=0)\end{cases}$	
det A	方阵的行列式	determinant of a square matrix	设 A 为方阵,则 det$A$ 表示 A 的行列式	A 的行列式亦可用 $\|A\|$ 表示
$\\|A\\|$	范数	norm	矩阵 A 的范数为 $\\|A\\|=(\mathrm{Tr}(AA^\dagger))^{\frac{1}{2}}$	范数有各种定义
$A_{m\times n}$ 或 $(a_{ij})_{m\times n}$	矩阵	matrix	$A_{m\times n}$ 表示一个 m 行 n 列的矩阵,$(a_{ij})_{m\times n}$ 表示 (i,j) 元素是 a_{ij} 的 m 行 n 列矩阵	
diag{…} 或[…]	对角矩阵	diagonal matrix	表示主对角线上元素为 $d_{11},d_{12},\cdots,d_{nn}$,其余元素全为零的方阵	
I 或 E	单位矩阵	unit matrix	表示主对角线上的元素都是 1,其他元素都是零的方阵,用 I 或 E 表示,称为单位矩阵	
A^{-1}	方阵 A 的逆	inverse of the square matrix A	设方阵 A 的行列式 $\|A\|\neq0$,则 $AA^{-1}=A^{-1}A=I$,其中 I 为单位方阵	
A^T 或 A'	A 的转置矩阵	transposed matrix of A	把矩阵 A 的行换成同序数的列,得到的新矩阵,称为 A 的转置矩阵	亦可表示成 $\tilde{A}$
$A\geqslant0$	非负矩阵	nonnegative matrix	实矩阵 A 中每个元素都是非负的	
$A>0$	正矩阵	positive matrix	实矩阵 A 中每个元素都是正的	
α^*	不增向量	nonincreasing vector	设 $\alpha=(a_1,a_2,\cdots,a_n)$ 是一个实向量. 若 $a_1^*,a_2^*,\cdots,a_n^*$ 是 $a_1,a_2,\cdots,a_n$ 的一个排列且满足 $a_1^*\geqslant a_2^*\geqslant\cdots\geqslant a_n^*$,则称 $\alpha\triangleq(a_1^*,a_2^*,\cdots,a_n^*)$ 是 α 的不增向量	

符号	中文名称	英文名称	意义或举例	备注
$\prec$	优于	major than	设$\boldsymbol{\alpha}=(a_1,a_2,\cdots,a_n)$,$\boldsymbol{\beta}=(b_1,b_2,\cdots,b_n)$是两个非负实向量,如果$a_1^*\leqslant b_1^*,\cdots,a_1^*+a_2^*+\cdots+a_{n-1}^*\leqslant b_1^*+b_2^*+\cdots+b_{n-1}^*$, $a_1+a_2+\cdots+a_n=b_1+b_2+\cdots+b_n$,则称$\boldsymbol{\beta}$优于$\boldsymbol{\alpha}$,记为$\boldsymbol{\alpha}\prec\boldsymbol{\beta}$	
Per A	积和式	formula of sum of products	A是$m\times n$复矩阵,$m\leqslant n$,$\mathrm{Per}\,A=\sum\limits_{\sigma}\prod\limits_{i=1}^{m}a_i\sigma(i)$称为$A$的积和式,其中$\sum$是对$\{1,2,\cdots,m\}$到$\{1,2,\cdots,n\}$的一切映射$\sigma$求和	
$\sigma(A)$	A的元素的和	sum of elements of A	表示矩阵A的所有元素之和	
$\rho(A)$	谱半径	spectral radius	设A为n阶复矩阵,$\lambda_1,\lambda_2,\cdots,\lambda_n$为其全部特征根,则$\rho(A)=\max\limits_{1\leqslant i\leqslant n}\lvert\lambda_i\rvert$称为$A$的谱半径	
(i,j)	(i,j)元素	(i,j) element	表示矩阵或行列式第i行第j列交叉位置上的元素	亦称(i,j)分量
A_{ij}	代数余子式	algebraic complement minor	在一个行列式中,(i,j)元素的代数余子式	
A^*	伴随矩阵	adjoint matrix	由n阶方阵A的所有元素的代数余子式A_{ij}为元素所构成的n阶方阵(A_{ij}置于第j行第i列交叉位置上)	亦可用$\bar{A}$或adj A表示
$\bar{A}$	增广矩阵	augmented matrix	在一个线性方程组的系数矩阵中,再在最后增加由常数项构成的列,所得到的矩阵	亦可用$\tilde{A}$表示
E_{ij}	矩阵单位	matrix unit	(i,j)元素是1,其余元素全是零的矩阵. 其中,$i=1,2,\cdots,m$;$j=1,2,\cdots,n$	多指方阵
TrA	方阵的迹	trace of a square matrix	方阵A的主对角线上所有元素之和	亦称追迹
rank (A)	矩阵的秩	rank of matrix	矩阵(不一定是方阵)A中不等于零的子式的最大阶数称为A的秩,零矩阵的秩规定是零	亦可用$r(A)$、"秩A"或"A秩"表示
$M_n(F)$,$F^{n\times n}$ $F_{n\times n}$,F_n	n阶全阵环	total matrix ring of order n	域F上全体n阶方阵对方阵的加法和乘法组成的环	更一般地,可把域F换成任意环R
$A\otimes B$	矩阵的直积	direct product of matrices	设$A=(a_{ij})_{m\times n}$,$B=(b_{ij})_{r\times s}$,则$mr\times ns$矩阵称为A与B的直积,记为$A\otimes B$	亦称 Kronecker 积.
$\dotplus$	方阵的直和	direct sum of a square matrix	设A为nk阶方阵.若A中表示成主对角线是k个n阶方阵$A_1,A_2,\cdots,A_k$,而其余块全为零的分块,则称A为$A_1,A_2,\cdots,A_k$的直和,记为 $A=A_1\dotplus A_2\dotplus\cdots\dotplus A_k$	
$\bar{A}$	A的复共轭矩阵	complex conjugate matrix of A	将复矩阵A的每个元素换成共轭复数所得矩阵记为$\bar{A}$,称为矩阵A的复共轭矩阵	
$\overline{A^+}$, $\overline{A^H}$	埃尔米特共轭矩阵	Hermitian conjugate matrix	矩阵A的复共轭矩阵$\bar{A}$的转置矩阵$\bar{A}'$,称为A的埃尔米特共轭矩阵	
A^+, A^H	埃尔米特矩阵	Hermitian matrix	若n阶矩阵A与它的转置共轭矩阵$\bar{A}'$相等,则A称为埃尔米特矩阵	
δ_{ik}	克罗内克δ	Kronecker's delta	$\delta_{ik}=\begin{cases}1 & (i=k),\\ 0 & (i\neq k)\end{cases}\quad(i,k=1,2,\cdots,n)$	
$R[x]$	多项式环	polynomial ring	系数属于环R、未知量(不定元)为x的全体多项式,对于多项式的普通加法和乘法组成的环	如果R有单位元1,则规定$x^0=1$
$R[x_1,x_2,\cdots,x_n]$	n元多项式环	n-ary polynomial ring	系数属于环R,未知量为$x_1,x_2,\cdots,x_n$(不相关不定元)的全体多项式,对于多元多项式的普通加法和乘法组成的环	如果环R有单位元1,则规定$x_i^0=1$,且$x_ix_j=x_jx_i$
deg $f(x)$	多项式的次数	degree of a polynomial	表示多项式$f(x)\neq0$中系数不为零的项中最高次项的次数	亦可用$\partial^\circ f(x)$表示
$\Phi_n(x)$	分圆多项式	cyclotomic polynomial	$\Phi_n(x)=\prod\limits_{i=1}^{\varphi(n)}(x-\xi_i)$称为$n$次分圆多项式,其中$\xi_1$,$\xi_2,\cdots,\xi_{\varphi(n)}$为$n$次原根	
$\sigma_1,\sigma_2,\cdots,\sigma_n$	初等对称多项式	elementary symmetrical polynomials	例如,x_1,x_2,x_3的初等对称多项式为:$\sigma_1=x_1+x_2+x_3$,$\sigma_2=x_1x_2+x_1x_3+x_2x_3$,$\sigma_3=x_1x_2x_3$	
$(f_1(x),\cdots,f_n(x))$	最高公因式	highest common factor	首系数为1且次数最高的公因式	亦称最大公因式
$[f_1(x),\cdots,f_n(x)]$	最低公倍式	least common multiple	首系数为1且次数最低的公倍式	亦称最小公倍式

符号	中文名称	英文名称	意义或举例	备注
$(f(x),g(x))=1$	互素	coprime	多项式 $f(x)$ 与 $g(x)$ 的最高公因式是1	
$(f_1(x),\cdots,f_n(x))=1$	两两互素	mutually prime	多项式 $f_1(x),f_2(x),\cdots,f_n(x)$ 中每两个都是互素的	
$F(x)$	有理分式域	rational traction field	域 F 上所有有理分式 $f(x)/g(x)(g(x)\neq 0)$ 关于有理分式的加法和乘法所组成的域	
$(a_1,a_2,\cdots,a_n)$	行向量	row vector	分量是 $a_1,a_2,\cdots,a_n$ 并排成一横行的 n 元向量	
$\begin{pmatrix}a_1\\a_2\\\vdots\\a_n\end{pmatrix}$	列向量	column vector	分量是 $a_1,a_2,\cdots,a_n$ 并排成一纵列的 n 元向量	
$\tau(i_1,i_2,\cdots,i_n)$	反序数	inverted sequence number	n 个数 $1,2,\cdots,n$ 的一个全排列 $i_1,i_2,\cdots,i_n$ 中反序个数的总和. 例如 $\tau(231)=2,\tau(321)=3$	亦称逆序数
$(i_1,i_2,\cdots,i_k)$	k 循环	k-cyclic(permutation)	即将 i_1 变为 i_2, i_2 变为 i_3, $\cdots$, i_k 变为 i_1, 而别的元素不动的置换	
sgn σ	置换的符号数	symbol number of permutation	设 σ 是一个置换,令 $\operatorname{sgn}\sigma=\begin{cases}+1 & (\sigma\text{ 是偶置换}),\\-1 & (\sigma\text{ 是奇置换})\end{cases}$	
(i,j)	对换	transposition	即将数码 i 变为 j, j 变为 i, 而别的数码不动的置换	
K^n	向量空间	vector space	以 K 为基域的 n 元向量的集合 K^n. 称为 K 上的向量空间或线性空间	当 $K=R$ 时记为 $\mathbb{R}^n$, 当 $K=\mathbb{C}$ 时记为 $\mathbb{C}^n$, 有时表示成 V
$\alpha\perp\beta$	正交向量	orthogonal vectors	内积为零的两个向量	
$\alpha\perp W$	向量与子空间正交	a vector cut a subspace orthogonally	欧氏空间中向量 α 与子空间 W 中每个向量都正交	亦可表示成 $(\alpha,W)=0$
$V_1\perp V_2$	正交子空间	orthogonal subspaces	V_1 与 V_2 是欧氏空间的两个子空间,若 V_1 中每个向量与 V_2 中每个向量都正交,则称 V_1 与 V_2 为正交子空间	
$W^{\perp}$	正交补	orthogonal complement	W 是欧氏空间 V 的一个子空间, $W^{\perp}$ 表示 V 中与 W 正交的一切向量所构成的子空间	
$\varphi\|W$	诱导变换	induced transformation	φ 是线性空间 V 的一个线性变换,子空间 W 对 φ 不变,则 φ 在 W 上的限制称为 φ 在 W 中的诱导变换	
$\leqslant$	子群	subgroup	$H\leqslant G$ 即 H 是群 G 的子群	亦可用 $<$ 表示子群或真子群
$\trianglelefteq$	正规子群	normal subgroup	$N\trianglelefteq G$ 即 N 是群 G 的正规子群	亦可用 $\triangleleft$ 表示正规子群或正规真子群
$\exp(G)$	有限群的指数	exponent of a finite group	设 G 是有限群,使 $a^n=1\ (\forall\, a\in G)$ 的最小正整数 n, 称为 G 的方次数	
$O_p(G)$	极大正规 p 子群	maximal normal p-subgroup	群 G 的极大正规子群且为 p 子群	
M^G	正规闭包	normal closure	群 G 的包含子集 M 的最小正规子群	
M_G	子集的核	core of subset	设 M 是群 G 的子集,则 G 的包含在 M 中的所有正规子群生成的子群称为 M 的核	
H char G	特征子群	characteristic subgroup	群 G 的在 G 的任意自同构下不变的子群	
$\mathrm{Syl}_p(G)$	西洛 p 子群	sylow p-subgroup	表示有限群 G 的一个西洛 p 子群,其中 p 是素数	
$S(G)$	基座	socle	群 G 的所有极小正规子群之积	
$\mathrm{Fit}(G)$	非廷子群	Fitting subgroup	群 G 的所有幂零正规子群之积	
$R_u(G)$	幂幺根基	unipotent radical	代数群 G 的最大连通正规幂幺子群	
$R(G)$	代数群的根基	radical of an algebraic group	代数群 G 的最大连通正规可解子群	
$\otimes,\times$	群的直积	direct product of groups	$G=G_1\times G_2\times\cdots\times G_n$ 或 $G=G_1\otimes G_2\otimes\cdots\otimes G_n$ 表示群 G 是群 $G_1,G_2,\cdots,G_n$ 的直积	群的直积有内外之分. 但在同构意义下可互相转化
$[X,Y]$	李括号	Lie bracket	$[X,Y]_p(f)=X_p(Yf)-Y_p(Xf)$	

符号	中文名称	英文名称	意义或举例	备注
$H \ltimes K$	半直积	semidirect product	$G/\overline{N} \cong F$，其中$\overline{N}$是与N同构的正规子群，$G = \overline{F}\overline{N}$，其中$\overline{F}$是与$F$同构的子群．$\overline{F} \cap \overline{N} = \{e\}$此时$G$称为$N$与$F$的半直积	
$\mathrm{N}_G(H)$	正规化子	normalizer	群G中所有可与子群H交换的元素组成的集合	定义子集S的正规化子为$\mathrm{N}_G(S)$
$\mathrm{C}_G(H)$	中心化子	centralizer	群G中所有与子群H的每个元素可交换的元素组成的集合	亦可表成$\mathrm{Z}_G(H)$
C_a	元素的中心化子	centralizer of an element	设a是群G的一个元素，则G中所有与a可交换的元素组成的集合	亦可记为$\mathrm{C}(a)$
$C(G)$	群的中心	center of a group	群G中与G的每个元素都可换的元素组成的集合	$C(G)$即$\mathrm{C}_G(G)$．亦可用$Z(G)$表示
$[a,b]$	换位子	commutator	群G中二元素a与b的换位子是指G中元素$a^{-1}b^{-1}ab$，即$[a,b] = a^{-1}b^{-1}ab$	换位子亦可定义为$[a,b] = aba^{-1}b^{-1}$
G'，(G,G)	换位子群	commutator group	由群G的一切换位子所生成的子群	亦称G的导出群或导群，并记为$D(G)$
$[A,B]$	A与B的换位子群	commutator subgroup of A and B	A,B是群G的两个子集．由所有换位子$[a,b]$($a \in A$，$b \in B$)所生成的子群	
$(G:H)$，$[G:H]$或$\|G:H\|$	子群的指数	index of a subgroup	子群H在群G中左(或右)陪集的个数．例如，$H = \{(1),(12)\} \subset S_3$，$(S_3:H) = 3$	$(G:H)$可能有限，也可能无限
$\Phi(G)$	弗拉蒂尼子群	Frattini subgroup	群G的所有极大子群的交	
$S(M)$，S_M	对称群	symmetric group	集合M的全体双射变换对变换乘法所组成的群，M可以是无限集	亦可表成$\mathrm{sym}(M)$
S_n	n次对称群	symmetric group of degree n	设$\|M\| = n$，则M上的对称群即M的全体双射变换对变换乘法组成的群，称为n次对称群	一般取$M = \{1,2,\cdots,n\}$
A_n	交错群	alternating group	n次对称群S_n中全体偶置换组成的群，称为n次交错群，简称交错群	亦称交代群
p^∞	p^∞型群	group of p^∞-type	$G = \bigcup\limits_{n=1}^{\infty} G_n$，其中$G_n$为所有$p^n$($p$是素数)次单位根对乘法组成的群．凡与$G$同构的群均称为$p^\infty$型群	亦称半循环群
$C(p^\infty)$	普吕费尔加群	prùfer additive group	设p是一固定素数，则所有形如a/p^n(n为任意正整数，a为任意整数)的有理数组成加群，它对于其子群$\mathbb{Z}$(整数加群)的商群(或称差群)称为普吕费尔加群	
$\|a\|$	元素的阶	order of the element	设a是群的元素．使$a^n = e$的最小正整数n，称为a的阶或周期．若这样的n不存在，则称a的阶是∞或0	亦可用$\circ(a)$表示
$\|G\|$	群的阶	order of a group	群G中所包含的元素的个数．例如，$\|S_3\| = 6$；整数加群$\mathbb{Z}$的阶为∞，即$\|\mathbb{Z}\| = \infty$	群G的阶也可记为$\mathrm{Ord}(G)$，而有限群G的阶也记为$[G:1]$
$\langle S \rangle$	由S生成的子群	generated subgroup by S	$\langle S \rangle$是群G中包含子集S的最小的子群，亦即G中包含S的所有子群的交．亦用(S)表示	当$S = \{a_1,a_2,\cdots,a_n\}$时，常记为$\langle a_1,a_2,\cdots,a_n \rangle$或$(a_1,a_2,\cdots,a_n)$
$\mathrm{Tor}G$	扭子群	torsion subgroup	群G的所有有限阶元素组成的子群，称为G的扭子群	亦称周期子群或挠子群
$\langle a \rangle$	循环群	cyclic group	由一个元素生成的群称为循环群．即$\langle a \rangle = \{\cdots a^{-2},a^{-1},e,a,a^2,\cdots\}$	亦可用(a)表示循环群
C_n	n阶循环群	cyclic group of order n	由一个阶为n的元素生成的循环群	
C_∞	无限循环群	infinite cyclic group	由一个阶为无限的元素生成的循环群记为C_∞	
$\rightarrowtail$	单同态	monomorphism	若φ是模A到B同态映射，而且又是单射时，记为$A \overset{\varphi}{\rightarrowtail} B$或$\varphi:A \rightarrowtail B$	多用在同调代数中模的同态上
$\twoheadrightarrow$	满同态	surjective homomorphism	若φ是模A到B的同态映射，而且又是满射时，记为$A \overset{\varphi}{\twoheadrightarrow} B$或$\varphi:A \twoheadrightarrow B$	多用在同调代数中模的同态上
$\leftrightarrow$，$\rightleftharpoons$	双射	bijection	表示集合M与$\overline{M}$间一个双射．例如，设$M = \{1,2,3,\cdots\}$，$\overline{M} = \{2,4,6,\cdots\}$，则$\varphi:n \leftrightarrow 2n$是双射	
$\simeq$	同态	homomorphism	$G \overset{\varphi}{\simeq} \overline{G}$表示$\varphi$是群$G$到群$\overline{G}$的一个同态．有时也简记为$G \simeq \overline{G}$	在环或其他代数系也有类似说法

符号	中文名称	英文名称	意义或举例	备注
$\cong$	同构	isomorphism	$G\cong\overline{G}$,表示群 G 与群 $\overline{G}$ 同构,即群 G 到群 G 存在一个保持运算的双射	对环、域、模等代数系的同构,亦用符号 $\cong$,$\backsimeq$ 或 $\simeq$ 表示同构
a^{φ}	元素的像	image of an element	φ是集合 A 到 B 的一个映射,$a\in A$.元素 a 在映射 φ 之下的像,一般用 $\varphi(a)$ 表示.亦用 a^{φ} 或 $a\varphi$ 表示	
G_{α}	稳定子群	stable subgroup	设 G 是 n 元集 $\Omega=\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$ 上的置换群,$\alpha\in\Omega$,则 $G_{\alpha}=\{g\mid g\in G,\alpha^g=\alpha\}$,即 G 中一切使 α 不动的置换组成的集合	G_{α} 是群 G 的一个子群
α^G	像的集合	set of image	设 G 是 n 元集 $\Omega=\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$ 上的置换群,$\alpha\in\Omega$,则 $\alpha^G=\{\alpha^g\mid g\in G\}$	α^G 是 Ω 的一个子集,且 $\|G\|=\|G_{\alpha}\|\cdot\|\alpha^G\|$
End G	自同态半群	endomorphism semi-group	群 G 的全体自同态对变换的乘法组成的半群	亦可记为 $E(G)$
Aut G	自同构群	automorphism group	群 G 的全体自同构对变换乘法组成的群	亦可简记为 $A(G)$
Inn$G$	内自同构群	inner automorphism group	G 是群,$a\in G,\tau_a:x\rightarrow axa^{-1}$ 是 G 的一个内自同构.G 的全体内自同构组成一个群,称为 G 的内自同构群	亦可简记为 $I(G)$.也把 axa^{-1} 写成 $a^{-1}xa$
Out(G)	外自同构群	group of outer automorphisms	群 G 的自同构群 Aut(G) 对于 G 的内自同构群 Inn(G) 的商群,称为 G 的外自同构群	
$R(G)$	右正则表示	right regular representation	G 为群,G 上一切置换 $\tau_g=\begin{pmatrix}x\\xg\end{pmatrix}$ $(g\in G)$ 组成的集合,称为群 G 的右正则表示	$R(G)$ 是 G 上对称群的子群
Hol$G$	全形	holomorph	$S(G)$ 为群 G 上的对称群,$R(G)$ 为 G 的右正则表示,$R(G)$ 在 $S(G)$ 中的正规化子称为群 G 的全形	
$\mathrm{GL}_n(F)$,$\mathrm{GL}(n,F)$	一般线性群	general linear group	域 F 上全体 n 阶可逆方阵对乘法组成的群,称为域 F 上的一般线性群,它与域 F 上的 n 维空间 V 的全体可逆线性变换组成的乘群 GL(V) 同构,故 GL(V) 亦称一般线性群	
$\mathrm{PGL}_n(F)$	射影一般线性群	projective general linear group	域 F 上 n 次一般线性群 $\mathrm{GL}_n(F)$ 关于其中心所得的商群,称为 F 上射影一般线性群	
$\mathrm{SL}_n(F)$,$\mathrm{SL}(n,F)$	特殊线性群	special linear group	表示域 F 上行列式等于 1 的全体 n 阶方阵对乘法组成的群	$\mathrm{SL}_n(F)$ 是 $\mathrm{GL}_n(F)$ 的正规子群
$\mathrm{PSL}_n(F)$	射影特殊线性群	projective special linear group	特殊线性群 $\mathrm{SL}_n(F)$ 关于其中心所得的商群,称为域 F 上的射影特殊线性群	
$O_n(F,S)$	正交群	orthogonal group	F 是特征不为 2 的域,S 是 F 上任意一个固定的 n 阶可逆对称矩阵,$O_n(F,S)=\{A\mid A\in F_{n\times n}$ 且 $A'SA=S\}$ 是一个群,称为 F 上(由 S 定义的)n 次正交群	
$O(n)$,O_n	实正交群	real orthogonal	由实数域上所有 n 阶正交方阵($A'=A^{-1}$)对乘法组成的群,称为 n 次实正交群	
$\mathrm{SO}(n)$	旋转群	rotation group	由实数域上所有行列式等于 1 的 n 阶正交方阵对乘法组成的群,称为 n 次旋转群	
$\mathrm{PO}_n(F,S)$	射影正交群	projective orthogonal group	正交群 $O_n(F,S)$ 关于其中心的商群	
$\mathrm{SP}_{2n}(F,J)$	辛群	symplectic group	J 是域 F 上 $2n$ 阶可逆交错矩阵 $F_{2n\times 2n}$ 中满足 $A'JA=J$ 的一切 A 组成的群,称为 F 上的 $2n$ 次辛群	
$\mathrm{PSP}_{2n}(F,J)$	射影辛群	projective symplectic group	辛群 $\mathrm{SP}_{2n}(F,J)$ 关于其中心的商群	
$U_n(F,K)$	酉群	unitary group	元素为复数的 n 阶酉矩阵的全体关于矩阵的乘法组成群,称为 n 维酉群	
SU	特殊酉群	special unitary group	$U(u)$ 中行列式等于 1 的所有矩阵形成 $U(u)$ 的正规子群,称为特殊酉群	
Spin	旋量群	spinor group	与 $\mathrm{SO}(n)$ 局部同构的单连通李群称为旋量群	
$\langle R,+,\cdot\rangle$	环	ring	非空集合 R 关于运算"+"与"·"组成的环记为$\langle R,+\cdot\rangle$,也常简记为 R	
$\leqslant$	子环	subring	$S\leqslant R$ 表示 S 是环 R 的子环	亦可用 $<$ 表示子环或真子环
Char R	特征(数)	character	R 为任意环.使 $na=0(\forall\ a\in R)$ 的最小正整数 n,称为 R 的特征.若这样的 n 不存在,称 R 的特征为 ∞ 或 0,例如,Char$Z_n=n$, Char$Z=\infty$	亦称特征数,环 R 的特征亦用 ch$R$ 表示

符号	中文名称	英文名称	意义或举例	备注
$U(R),R^*$	单位群	unit group	R是有单位元的环,R的全体单位(即可逆元)对R的乘法组成群,称为R的单位群.例如,整数环$\mathbf{Z}$的单位群为$U(\mathbf{Z})=\{1,-1\}$	R的单位群亦称R的乘群
R^0	逆环	inverse ring	R为环.如果保持R的加法不变,而乘法改为$a\circ b=ba$,则R对于原加法和新乘法。也组成环,称为R的逆环	亦称反环,并记为R^{op}
$\mathbf{Z}[i]$	高斯整环	Gaussian integral domain	由一切复数$a+bi$ $(a,b\in\mathbf{Z})$所组成的数环	
$R[G]$	群环	group ring	设R是有单位元的环,G为群,一切有限和$\sum a_ix_i(a_i\in R,x_i\in G)$关于其(类似于多项式的)加法与乘法组成的环	亦可记为$R(G),RG$或GR
$F(G)$	群代数	group algebra	域F和群G构成的群环$F[G]$,再加上F中元素与有限和$\sum a_ix_i(a_i\in F,x_i\in G)$的乘法而得到的$F$上的代数	
$J(R)$	雅各布森根	Jacobson radical	环R的所有本原理想的交,称为R的雅各布森根.当R无本原理想时,规定:$J(R)=R$	亦简称J根,有多种定义方法
$\trianglelefteq$	理想	ideal	$I\trianglelefteq R$表示I是环R的理想	亦可用$\triangle$表示理想或真理想
$\langle a\rangle$	主理想	principal ideal	环中包含元素a的最小理想	亦可用(a)表示
$\oplus$或$\dotplus$	环的直和	direct sum of rings	$R=R_1\oplus R_2\oplus\cdots\oplus R_n$或$R=R_1\dotplus R_2\dotplus\cdots\dotplus R_n$,即环$R$是$R_1,R_2,\cdots,R_n$的直和	对于加群的直积也常称为直和;又子空间的直积,都常用$\oplus$或$\dotplus$表示
$\sqrt{A}$	理想的根	radical of an ideal	A为交换环R的理想.$\sqrt{A}=\{a\mid a\in R,\exists\ n$使$a^n\in A\}$($n$与$a$有关),称为理想$A$的根	亦称理想A的根基
$\langle S\rangle$	由S生成的理想	generated ideal by S	S是环R的一个子集,$\langle S\rangle$是R中包含S的最小理想,亦即R中包含S的所有理想的交	当$S=\{a_1,a_2,\cdots,a_n\}$时,常记为$\langle a_1,a_2,\cdots,a_n\rangle$或$(a_1,a_2,\cdots,a_n)$
AB	理想的积	product of ideals	A,B是环R的理想,则一切有限和$\sum a_ib_i(a_i\in A,b_i\in B)$组成$R$的一个理想,称为理想$A$与$B$的积	AB是由一切元素$ab(a\in A,b\in B)$生成的理想
$A:B$	理想的商	quotient of ideals	设A,B是交换环R的理想,则R中满足$xB\subseteq A$的一切元素x组成R的理想,称为A与B的商	
$O:B$	零化理想	annihilating ideal	设B是交换环R的理想,则R中满足$xB=0$的一切元素x组成的理想,称为B的零化理想	当R为非交换时,$O:B$是R的左理想
$l(S)$,ann S_l	左零化子	left annihilator	环R中使$rS=0$的一切r组成的集合	$l(S)$是R的左理想
$r(S)$,ann S_r	右零化子	right annihilator	环R中使$Sr=0$的一切r组成的集合	$r(S)$是R的右理想
N_K	克德根	Köthe radical	环R的最大幂零元理想,称为R的克德根,简称K根	
N_Q	近似诣零根	quasi-nil radical	环R的全部近似诣零单边理想之和,称为R的近似诣零根	
N_L	林文茨基根	Livitzki radical	环R的惟一最大局部幂零理想称为R的林文茨基根	
N_{BM}	布朗-麦柯根	Brown-Mccoy radical	环R的最大g正则理想,称为R的布朗-麦柯根	
$F(\alpha)$	单扩张	simple extension	包含域F和元素α的最小扩域	亦称单扩域
$F(S)$	域的扩张	extension of a field	E是域F的扩域,S是E的一个子集,E中包含F和S的最小域记为$F(S)$,它是域F的扩张	当$S=\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$时,则$F(S)$记为$F(\alpha_1,\alpha_2,\cdots,\alpha_n)$
$(E:F)$,$[E:F]$	扩域次数	degree of an extended field	E是域F的扩域,则E是域F上的向量空间.E在F上的维数称为扩域的次数或扩张次数	$(E:F)$可能有限,也可能无限
$A(E\|F)$	E在F上的伽罗瓦群	Galois group of E over F	F是域E的子域,$A(E\|F)$是E的使F的每个元素不动的全体自同构组成的群	
$E(G_1)$	子群G_1所属的域	field belong to subgroup	E是域F的扩域,又$G=A(E\|F)\geqslant G_1$,E中所有对于G_1中任一元都不动的元是E的子域,称为子群G_1所属的域	$F\subseteq E(G_1)\leqslant E$
$G(E_1)$	子域E_1所属的群	group belong to subfield	假设同上,又E_1是E的子域且$F\subseteq E_1\subseteq E$.则$G$中所有不使$E_1$中任意元变动的元素之集是$G$的子群,称为子域$E_1$所属的群	

符 号	中文名称	英文名称	意义或举例	备 注
F_q, GF(q)	有限域	finite field	F_q 或 GF(q)表示元素个数为 q 的有限域	元素个数相同的有限域都同构
$\mathbf{Q}_p$	p 进数域	p-adic number field	表示有理数域在 p 进赋值下的完备化域	p 为素数
$\mathbf{Z}_p$	p 进整数环	ring of p-adic integers	全体 p 进整数组成的环,称为 p 进整数环	p 为素数
$K[[\ \ ]]$	形式幂级数环	formal power series ring	$K[[x_1, x_2, \cdots, x_n]]$表示系数在域 K 中的形式幂级数环	亦可表示成 $R\{x_1, x_2, \cdots, x_n\}$
$G_rU(A)$	分次单位群	graded unit group	G 为群,$U(A)$ 是 G 分次代数 $A = \bigoplus_{g\in G} A_g$ 的单位群. A 的一切分次单位组成 $U(A)$ 的一个子群	
GS(V)	半线性变换群	semilinear transformation group	V 是域 F 上的向量空间,V 的一切非奇异半线性变换组成群,称为半线性变换群	
$J_G(M)$	雅各布森分次根	Jacobson graded radical	R 为 G 分次环,M 为分次 R 模. M 的一切分次极大模的交,称为 M 的雅各布森分次根	
δ	导子	derivation	环 R 的导子,即 R 的满足 $\delta(a+b)=\delta a+\delta b$ 与 $\delta(ab)=(\delta a)b+a(\delta b)$ 的变换 δ	
$D(A)$	A 上微分算子环	ring of differential operators over A	称 $\bigcup_{i=0}^{\infty} D^i(A)$ 为 A 上线性微分算子环	
deg A	代数 A 的次数	degree of algebra A	设 A 是域 F 上中心单代数,且$(A:F)=m^2$,则称 m 为 A 的次数	
Ind$A$	舒尔指数	Schur index	A 是域 F 上有限维中心单代数,且 $A \cong M_n(D)$,其中 D 是 F 上可除代数,称 deg$D$ 为 A 的舒尔指数	
Bsi$A$	次理想	subideal	设 B 是代数 A 的一个子代数,若有 $B=B_0 \subseteq \cdots \subseteq B_n = A$,其中 B_i 是 B_{i+1} 的理想,则称 B 是 A 的次理想	
$\triangle_T$	T 理想	T-ideal	设 I 是代数 A 的一个理想. 如果对 A 的每个自同态 φ 均有 $\varphi(I) \subseteq I$,则称 I 为 A 的 T 理想	
$S^{-1}R$	分式环	ring of fractions	设 R 是有单位元的交换环,S 是 R 的乘闭子集. 则一切 $a/s(\forall\, a \in R, s \in S)$ 关于分式的加法和乘法组成环,称为 R 关于 S 的分式环	
$P^{(n)}$	符号幂	symbolic power	设 P 是有单位元的交换环 R 的素理想,$S_P = R\backslash P$. 称 $S_P^{-1}P^n$ 在 R 中的收缩理想为 P 的 n 次符号幂	
(x,y,z)	结合子	associator	称 $(xy)z - x(yz)$ 为非结合代数中三个元素 x,y,z 的结合子	
Der(R)	导子李环	Lie ring of derivations	结合环 R 的导子在加法与乘法$[\delta_1,\delta_2]=\delta_1\delta_2-\delta_2\delta_1$ 之下组成的李环,称为导子李环	
Corad(C)	余代数的余根	coradical of coalgebra	余代数 C 的所有单子余代数的和,称为 C 的余根	
$l(K\|F)$	F 共轭映射数	number of F-conjugate mapping	设 Ω 是域 F 的扩域 K 的代数闭包,则 K 到 Ω 的一切 F 共轭映射的个数记为 $l(K\|F)$	
tr. $\deg_F K$	超越次数	transcendence degree	域 F 的扩域 K 的超越基的基数称为 K 在 F 上的超越次数	
$N_F^K(\alpha)$	α 的范	norm of α	K 是域 F 的有限次扩域,Ω 是 F 的含 K 的代数闭包;又 $\sigma_1,\sigma_2,\cdots,\sigma_m$ 为 K 到 Ω 的一切互异的 F 共轭映射,则 $N_F^K(\alpha) = \left(\prod_{j=1}^{m}\sigma_j(\alpha)\right)^{[K:F]_i}$ 称为 K 中元 α 的范	
$T_F^K(\alpha)$	α 的迹	trace of α	K 是域 F 的有限次扩域,Ω 是 F 的含 K 的代数闭包;又 $\sigma_1,\sigma_2,\cdots,\sigma_m$ 为 K 到 Ω 的一切互异的 F 共轭映射,则 $T_F^K(\alpha) = [K:F]_i \cdot \sum_{j=1}^{m}\sigma_j(\alpha)$ 称为 K 中元 α 的迹	
X_F	正锥集	set of positive cone	X_F 表示实域 F 的全部正锥组成的集合	
$X_F(T)$	序空间	space of orderings	T 是实域 F 的一个亚正锥,$X_F(T)$ 表示 F 上所有包含 T 的正锥所组成的集合,称为亚序域(F,T)的序空间	
$H(F)$	实全纯环	real holomorphic ring	实域 F 的所有实赋值环的交是 F 的一个子环,称为 F 的实全纯环	
(F,φ)	赋值域	valued field	带有赋值 φ 的域 F,称为赋值域	带有赋值环 B 的域 F 记为(F,B)
M_R	右 R 模	right R-module	R 是有单位元的环,M_R 是右 R 模,即作用乘法为 $ar(a \in M, r \in R)$	类似地有左 R 模${}_RM$

符号	中文名称	英文名称	意义或举例	备注
$\hookrightarrow$	子模	submodule	$A \hookrightarrow M$ 表示 A 是模 M 的一个子模	
$\hookrightarrow$	小子模	small submodule	设 A 是模 M 的一个子模.如果对 M 的任意子模 Z 有 $A+Z=M$ 必有 $Z=M$,则称 A 为 M 的小子模,记为 $A \hookrightarrow M$	即只有 M 才使 $A+M=M$ 的子模 A 称为小子模
$\overset{*}{\hookrightarrow}$	大子模	large submodule	设 A 为模 M 的子模,若对 M 的任意子模 Z 有 $A \cap Z=0$ 必有 $Z=0$,则称 A 为 M 的大子模,记为 $A \overset{*}{\hookrightarrow} M$	即只有 $\{0\}$ 使 $A \cap \{0\}=0$ 的子模 A 称为大子模
$\mathrm{Si}(M)$	奇异子模	singular submodule	设 M 为右 R 模,M 中所有使 $r_r(m) \overset{*}{\hookrightarrow} R_R$ 的 m 组成的集是 M 的子模,称为奇异子模,其中 $r_R(m)=\{r \mid r \in R, mr=0\}$	
$\mathrm{ann}_R\, x$	阶理想	order ideal	设 R 是有1环,M 是左 R 模,$x \in M$,记 $\mathrm{ann}_R x=\{a \in R \mid ax=0\}$,称为 x 在 R 中的阶理想	亦称为 x 在 R 中的零化子.记为 $(0 : x)$
M^+	特征模	character module	M 是左 R 模,$M^+=\mathrm{Hom}_Z(M, Q/Z)$ 对于 $(f \circ r)(x)=f(rx)(f \in M^+, r \in R, x \in M)$ 组成右 R 模,称为 M 的特征模	
$\mathrm{G.dim}(M)$	戈迪维数	Goldie dimension	若 R 模 M 有子模 $U_1, U_2, \cdots, U_n$ 使 $\sum\limits_{i=1}^{n} U_i$ 为直和且为 M 的本质子模,则称 n 为 M 的戈迪维数	
R-Mod	R 模范畴	category of R-modules	所有左 R 模构成的范畴,称为左 R 模范畴	
$H^n(X)$	上同调模	cohomology modules	令 $X: \cdots \to X^{n-1} \xrightarrow{d^{n-1}} X^n \xrightarrow{d^n} X^{n+1} \to \cdots$ 是环 R 上的复形,$H^n(X)=\ker d^n/\mathrm{Im} d^{n-1}$,称为 X 的上同调模	
$\mathrm{Ext}_R^n(M,-)$	函子 Ext	functor Ext	设 M 是右 R 模,用 $\mathrm{Ext}_R^n(M,-)$ 表示 $\mathrm{Hom}_R(M,-)$ 的右导出函子	
$\mathrm{Tor}_n^R(M,-)$	函子 Tor	functor Tor	设 M 是右 R 模,用 $\mathrm{Tor}_n^R(M,-)$ 表示 $M \otimes_R -$ 的左导出函子	
$\mathrm{l \cdot Pd}_R M$	左投射维数	left projective dimension	表示 M 为左 R 模,M 的左投射维数	亦称左同调维数,记为 $\mathrm{l \cdot dh}_R N$
$\mathrm{r \cdot pd}_R N$	右投射维数	right projective dimension	表示 N 为右 R 模,N 的右投射维数	亦称右同调维数,记为 $\mathrm{r.dh}_R N$
$\mathrm{l.gl.dim}\, R$	左整体维数	left global dimension	环 R 的左整体维数 $\mathrm{l.gl.dim}\, R=\sup\{\mathrm{l.pd}_R M \mid M \in \mu_R\}$	
$\mathrm{r.gl.dim}\, R$	右整体维数	right global dimension	环 R 的右整体维数 $\mathrm{r.gl.dim}\, R=\sup\{\mathrm{r.pd}_R M \mid M \in \mu_R\}$	
$\mathrm{l.Id}_R M$	左内射维数	left injective dimension	表示左 R 模 M 的左内射维数	
$\mathrm{r.Id}_R N$	右内射维数	right injective dimension	表示右 R 模 N 的右内射维数	
$\mathrm{l.Fd}_R M$	左平坦维数	left flat dimension	表示左 R 模 $M \neq 0$ 的左平坦维数	亦称弱左同调维数,记为 $\mathrm{w.l.dh}_R M$
$\mathrm{r.Fd}_r N$	右平坦维数	right flat dimension	表示右 R 模 $N \neq 0$ 的右平坦维数	亦称弱右同调维数记为 $\mathrm{w.r.dh}_R N$
$M_1 * M_2 * \cdots * M_n$	双积	biproduct	设 M 及 $M_1, M_2, \cdots, M_n$ 为 R 模.若有模同态 $\sigma_i: M_i \to M$ 与 $\pi_j: M \to M_j$ 满足 $\pi_j \sigma_i=\delta_{ji}$ 与 $\sum \sigma_i \pi_i=1_M$,则称 $\pi_j M$ 是模 $M_1, M_2, \cdots, M_n$ 的双积	
$\mathrm{Obj}(K)$	对象类	class of objects	K 是一个范畴,K 的所有对象构成的类称为 K 的对象类	
$\mathrm{Mor}_K(A,B)$	(态)射集	set of morphisms	A, B 是范畴 K 的两个对象.由 A 与 B 所决定的一个集合称为 A 与 B 的(态)射集	亦称为由 A 到 B 的射或态射
$\mathrm{Dom}(\alpha)$	(态)射的域	domain of a morphism	表示在范畴中,设 $\alpha \in \mathrm{Mor}_K(A,B)$,则称 A 为(态)射 α 的域	
$\mathrm{Cod}(\alpha)$	(态)射的上域	codomain of a morphism	在范畴中,当 $\alpha \in \mathrm{Mor}_K(A,B)$ 时,称 B 为(态)射 α 的上域	
$\mathrm{rad}(M)$	模的根	radical of a module	表示模 M 的所有极大子模的交	亦即 M 的所有小子模的和
$Soc\,(M)$	模的基座	socle of module	表示模 M 的所有极小子模的和	亦即 M 的所有大子模的交
$\ker \varphi$	核	kernel	φ 是环 R 模 A 到 B 的一个同态映射,称 B 中零元素的全体逆象 $\varphi^{-1}(0)$ 为 φ 的核	对群、环等代数系也有类似概念

符号	中文名称	英文名称	意义或举例	备注
Coker φ	上核	cokernel	φ是环R模A到B的一个同态映射，商模$B/\mathrm{Im}\varphi$称为φ的上核	亦称余核
Coim φ	上象	coimage	φ是环R模A到B的一个同态映射，商模$A/\mathrm{Ker}\,\varphi$称为φ的上像	亦称余像
M/N	商空间	quotient space	表示两代数系M,N的商空间	
dim V	维数	dimension	表示线性空间V的维数	
V^*	对偶空间	dual space	域F上线性空间V的所有线性函数组成F上的线性空间，称为V的对偶空间	V^*即$\mathrm{Hom}_F(V,F)$
$W(A)$	矩阵的数值域	numerical range of a matrix	$A\in C^{n\times n}$，称$W(A)=\{x^*Ax\mid x\in C^n, x^*x=1\}$为$A$的数值域	
$r(A)$	矩阵的数值半径	numerical radius of a matrix	$A\in C^{n\times n}$，称$\max\limits_{Z\in W(A)}\|Z\|$为$A$的数值半径	
V_{λ_0}	特征子空间	characteristic subspace	设σ是线性空间V的一个线性变换，λ_0是σ的一个特征值，则对应于λ_0的全体特征向量和零向量组成的子空间称为特征子空间	
$T(G,x)$	对称化算子	symmetrization operator	张量空间$T_p^0(E)$或$T_p^0(E)$的线性变换$S_p=\sum\limits_{\sigma\in G_p}\sigma$称为对称化算子，其中$G_p$为置换群	
$V_x(G)$	张量对称类	symmetric class of tensors	设$\overset{m}{\otimes}V$是张量空间，x是群G的不可约特征标，$T(G,x)$是对称化算子，则称$\mathrm{Im}T(G,x)$为关于G和x的张量对称类	
Inex $V_\chi(G)$	张量对称类的指标	index of symmetric class of tensor	表示张量对称类$V_\chi(G)$的指标	
$d_f^G(A)$	广义矩阵函数	generalized matrix function	设$A=(a_{ij})$为m阶复方阵，G为S_m的子群，f是G到C的任一函数，则称$d_f^G(A)=\sum\limits_{\sigma\in G}f(\sigma)\prod\limits_{t=1}^{m}a_{t\sigma(t)}$为广义矩阵函数	
$E(V)$	外代数	exterior algebra	设V为域$K(\mathrm{char}K\neq 2)$上向量空间，$\overset{m}{\wedge}V$为K上的格拉斯曼空间，则直和$\overset{0}{\wedge}V\oplus\overset{1}{\wedge}V\oplus\cdots\oplus\overset{n}{\wedge}V$可组成$K$上代数，称为$V$上的外代数	亦称格拉斯曼代数
$\vee E$	对称代数	symmetric algebra	设E是域$K(\mathrm{char}k=0)$上的向量空间，$\vee^p$是E的p次对称幂，则$\vee E=\bigoplus\limits_{p=0}^{\infty}\vee^pE$可组成$K$上交换代数，称为$E$上的对称代数	
S_V	对合S_V	involution S_V	设V是域K上向量空间，则包含映射$j:V\to C\ell^P$在$C_V\to C\ell^P$的代数开拓是一个对合，其中$C\ell^P$是V的克利福德代数C_V的反代数	
$\widehat{\oplus}$	正交直和	orthogonal direct sum	设$U_1,U_2,\cdots,U_m$是V的向量子空间，若它们两两正交且V为其直和，则记为$V=U_1\widehat{\oplus}\cdots\widehat{\oplus}U_m$，称$V$为$U_i$的正交直和	
$\uplus$	格-并	lattice-union	A$\uplus$B表示两个理想A,B的格-并	
$\boldsymbol{C}^0$	对偶范畴	dual category	由范畴$\boldsymbol{C}$作出的新范畴$\boldsymbol{C}^0$：$\boldsymbol{C}^0$的对象类即$\boldsymbol{C}$的对象类，定义$\mathrm{Hom}_{C^0}(A^0B^0)=\mathrm{Hom}_C(B,A)$，并规定$f^0g^0=(gf)^0$，称$\boldsymbol{C}^0$为$\boldsymbol{C}$之对偶范畴	
Set	集范畴	category of sets	以一切集合为对象，以集合映射为态射的范畴	
Top	拓扑空间范畴	category of topological spaces	以一切拓扑空间为对象，以连续映射为态射的范畴	亦可表示成$\mathscr{T}$
Group	群范畴	category of groups	以一切群作对象，以群同态作态射的范畴	亦可表示成$\mathscr{G}$
AG	阿贝尔群范畴	category of Abelian groups	以一切阿贝尔群作对象，以阿贝尔群同态作态射的范畴	
Ring	环范畴	category of rings	以一切环作对象，以环同态作态射的范畴	亦可表示成μ_R
$\prod\limits_{\lambda\in\Lambda}\boldsymbol{C}_\lambda$	积范畴	product category	$\{\boldsymbol{C}_\lambda\}(\lambda\in\Lambda)$为一个范畴集合．由它们所作出的新范畴$\prod\limits_{\lambda\in\Lambda}\boldsymbol{C}_\lambda$为$\{\boldsymbol{C}_\lambda\}$的积范畴	
$\coprod\limits_{\lambda\in\Lambda}A_\lambda$	上积	coproduct	$\{A_\lambda\}(\lambda\in\Lambda)$为范畴$\boldsymbol{C}$的一个对象集．若对象$B\in\boldsymbol{C}$与一态射集具有泛性质，则称$B$为$\{A_\lambda\}$的上积	

符 号	中文名称	英文名称	意义或举例	备 注
IBN	IBN 环	IBN ring	R 为环.如果每个有限生成的 R 模的任二基中元素个数必相等,则称 R 为 IBN 环	
$(\mathscr{C},\perp)$	带积范畴	category with product	规定映射 $\perp:\mathscr{C}\times\mathscr{C}\to\mathscr{C}$ 的范畴 $\mathscr{C}$ 称为带积范畴	
ΦF	纤维范畴	fibre category	$(\mathscr{C},\perp)$ 与 $(\mathscr{D},T)$ 为带积范畴,$F:\mathscr{C}\to\mathscr{D}$ 为保积函子.由此定义的新范畴 ΦF(对象类为 $\{(M,N,\alpha)\mid M,N\in\mathscr{C},\alpha:F(M)\cong F(N)\}$ 称为 $\mathscr{C}$ 与 $\mathscr{D}$ 的纤维范畴)	
$\mathrm{gl}(V)$	一般线性李代数	general linear lie algebra	$\mathrm{gl}(V)$ 表示域上 n 维空间 V 的所有线性变换在运算 $[A,B]=AB-BA$ 下组成的 n^2 维李代数,称为一般线性李代数	
$n(P)$	偏序集的阶	order of poset	偏序集 P 的基数称为 P 的阶	
$l(P)$	偏序集的长	length of poset	偏序集 P 中链的长的最小上界称为 P 的长	
$\mathrm{Sup}\ X$	上确界	supremum	偏序集的子集 X 的上确界	亦称最小上界.记为 $\vee X$ 或 l.u.b.X
$\inf X$	下确界	infimum	偏序集的子集 X 的下确界	亦称最大下界.记为 $\wedge X$ 或 g.l.b.X
$(L;\leqslant)$	格	lattice	若偏序集 L 的任二元素均有上确界和下确界,则称 L 为格	
$\Phi(L)$	弗拉梯尼子格	Frattini sublattice	表示格 L 的弗拉梯尼子格	
a^+	a 的正部	positive part of a	a 是格群的一个元素,$a^+=a\vee 0$ 称为 a 的正部	
a^-	a 的负部	negative part of a	a 是格群的一个元素,$a^-=(-a)\vee 0$ 称为 a 的负部	
$X^\perp$	极	polar	X 是格群 G 的子集,$X^\perp=\{y\in G\mid \lvert y\rvert\wedge\lvert x\rvert=0,\forall x\in X\}$,称为 X 的极	
$J\perp K$	独立 l 理想	independent l-ideal	格序群的 l 理想 J,K 若有 $J\wedge K=0$,则称 J 和 K 是独立的	
$R(G)$	康莱德根	Conrad radical	格序群 G 的一切本质性值的交是一个 l 理想,称为 G 的康莱德根	
R^+	偏序环的序	order of po-ring	R 是偏序环,$R^+=\{r\in R\mid r\geqslant 0\}$,称为 R 的序	亦称 R 的正锥
BCK	BCK 代数	BCK-algebra	一种有序代数系统	
BCI	BCI 代数	BCI-algebra	一种较 BCK 代数广泛的代数结构	
$(X;*,0)$	双 B 代数	two B-algebra	表示 BCK 代数或 BCI 代数,二者合称双 B 代数	
A^*	稳定子	stabilizer	A 是 BCK 代数 X 的子集,$A^*=\{x\in X\mid x*a=x$ 且 $a*x=a,\forall a\in A\}$,称为 A 的稳定子	
(X,O_X)	环式空间	ringed space	带有一个环层 O_X 的拓扑空间 X,称为环式空间	
$\chi(O_X)$	欧拉-庞加莱特征标	Euler-Poincaré characteristic	n 维完备簇 X 的欧拉-庞加莱的特征标定义为 $\chi(O_X)=\sum_{i=0}^{n}(-1)^i\dim_k H^i(X,O_Z)$	
$K(X)$	小平维数	Kodaira dimension	X 是 n 维完备代数簇.在 X 利用归纳法定义的维数 $K(X)$ 称为小平维数	
$R(X)$	典范环	canonical ring	X 为光滑射影族,ω_E 为其典范层,X 的典范环为 $R(X)=\bigoplus_{n\geqslant 0}H^0(X,\omega_X^{\otimes n})$	
$\mathrm{Pic}(X)$	皮卡群	Picard group	环式空间 (X,O_X) 的可逆层的同构类组成的群(运算由可逆层的张量积所诱导),称为 X 的皮卡群	
$\mathrm{Pic}^0(X)$	皮卡簇	Picard variety	X 是代数闭域 K 上的射影光滑代数簇,$\mathrm{Pic}(X)$ 中包含 O 的分支是一个射影概形,它的既约结构是一个阿贝尔簇,称为 X 的皮卡簇	
$\mathrm{Alb}(Z)$	阿尔班尼斯簇	Albanese variety	X 是射影光滑代数簇.X 的皮卡簇的对偶阿贝尔簇称为 X 的阿尔班尼斯簇	
$G_{n,m}$	格拉斯曼簇	Grassmannian variety	一个 n 维线性空间的所有 m 维线性子空间的集合称为一个格拉斯曼簇	亦称格拉斯曼流形或格拉斯曼空间
$\mathrm{Flag}(n_1,n_2,\cdots,n_r)$	旗簇	flag variety	V 是 n 维向量空间,$n=n_1>n_2>\cdots>n_r>0$.则 V 的所有由子空间组成的指标为 $(n_1,n_2,\cdots,n_r)$ 的旗的集合,称为一个旗簇	
$\times$	叉积	cross product	a,b 的叉积等于 a,b 的对称差的补运算,即 $a\times b=(a\triangle b)'$	这里 $a,b\in B$,B 称为布尔集

符号	中文名称	英文名称	意义或举例	备注
c	胞腔度	cellularity	cA$=\sup\{\lvert x\rvert \mid x$ 是其中的一个两两不相交的族}. 称为布尔代数 A 的胞腔度	
sat A	浸润度	saturation	sat$A=\min\{u \mid u$ 是基数且对 A 的每个两两不相交的族 x 有 $\lvert x\rvert<u\}$ 表示 A 的浸润度，它是一个正则基数，式中 $\lvert x\rvert$ 表示 x 的基数	
π	稠密度	density	$\pi B=\min\{\lvert x\rvert \mid x\subseteq B$ 在 B 中稠密}表示 X 在布尔代数 B 中的稠密度	
Id	理想	ideal	Id(B)表示布尔代数 B 中的全体理想	布尔代数 B 中的每个理想记为 I，有限集的理想记为 fin
Sub	子代数	subalgebra	Sub$A$ 表示无限布尔代数 A 的一切子代数所构成的集合	sub(B)表示布尔代数 B 的子代数所构成的格
Ult	超滤子	ultrafilter	Ult$A$ 表示无限布尔代数 A 的超滤子的全体	
Filt	滤子	filter	Filt$A$ 表示无限布尔代数 A 的一切滤子所构成的集合	
Σ	最小上界	least upper bound	$\Sigma^B M$ 表示 M 在布尔代数 B 中的最小上界，其中 M 是 B 的子集	
clop	闭开代数	clopen algebra	拓扑空间 X 的所有闭开集，用 clop$X$ 表示，构成 X 上的集合代数称为 X 的闭开代数	
RO( )	正则开代数	regular open algebra	RO(x)$=\{u \mid u\subseteq X$ 且 $r(u)=u\}$，其中 $r(u)=\mathrm{int}(\mathrm{cl}(u))$是 u 的正则化	
Bai	贝尔代数	Baire algebra	Bai$X=\{a\subseteq X \mid a$ 有贝尔性质}，其中 a 是拓扑空间 X 的子集，存在 X 的一个开集 u，使对称差 $a\triangle u$ 是贫集	
A↾a	相对代数	relative algebra	$A\restriction a=\{x \mid x\in A$ 且 $x\leqslant a\}$表示 A 关于 a 的相对代数. 式中 A 是布尔代数，且 $a\in A$	亦称因子代数
pred(t)	前趋集合	predecessor set	偏序集$(T,\leqslant_T)$是一棵树，且所有的 $t\in T$，集合 pred(t)是由$<_T$ 决定的一个良序集合	
Tor	挠积	torsion product	$\mathrm{Tor}_n(M,N)$是 M 和 N 的挠积	
Ext	扩张	extension	$\mathrm{Ext}^n(M,N)$是 M，N 的扩张	

分析学(analysis)

符号	中文名称	英文名称	意义或举例	备注
(a,b)	开区间	open interval	表示 a 与 b 之间(不包括端点 a 与端点 b)的一切实数组成的集合	亦可用$]a,b[$表示
$[a,b]$	闭区间	closed interval	表示 a 与 b 之间(包括端点 a 与端点 b)的一切实数组成的集合	
$(a,b]$	左半开区间	left half open interval	表示 a 与 b 之间(不包括端点 a 但包括端点 b)的一切实数组成的集合	亦可用$]a,b]$表示
$[a,b)$	右半开区间	right half open interval	表示 a 与 b 之间(包括端点 a 但不包括端点 b)的一切实数组成的集合	亦可用$[a,b[$表示
e^x 或 $\exp x$	指数函数	exponential function	表示以 e 为底，以 x 为指数的函数，可写成 $y=\mathrm{e}^x$ 或 $y=\exp x$	在同一场合中，只用其中一种符号
e	超越数	transcendental number	$\mathrm{e}=\lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n=2.718\ 281\ 828\ 459\cdots$	通常作为自然对数的底
$\log_a x$	对数函数	logarithmic function	表示以 a 为底，自变量为 x 的对数函数，可写成 $y=\log_a x$	
$\ln x$	自然对数	natural logarithm	表示以 e 为底，自变量为 x 的对数函数	
$\lg x$	常用对数	common logarithm	表示以 10 为底，自变量为 x 的对数函数	
$\mathrm{lb}\, x$	2 为底的对数	logarithm to the base 2	表示以 2 为底，自变量为 x 的对数函数	亦可记为 $\log_2 x$
$\mathrm{sh}\, x$ 或 $\sinh x$	双曲正弦	hyperbolic sine	$\mathrm{sh}\, x=\dfrac{\mathrm{e}^x-\mathrm{e}^{-x}}{2}$	

符号	中文名称	英文名称	意义或举例	备注
ch x 或 cosh x	双曲余弦	hyperbolic cosine	$\mathrm{ch}\,x=\frac{e^x+e^{-x}}{2}$	
th x 或 tanh x	双曲正切	hyperbolic tangent	$\mathrm{th}\,x=\frac{\mathrm{sh}\,x}{\mathrm{ch}\,x}=\frac{e^x-e^{-x}}{e^x+e^{-x}}$	
coth x	双曲余切	hyperbolic cotangent	$\mathrm{coth}\,x=\frac{\mathrm{ch}\,x}{\mathrm{sh}\,x}=\frac{e^x+e^{-x}}{e^x-e^{-x}}$	
sech x	双曲正割	hyperbolic secant	$\mathrm{sech}\,x=\frac{1}{\mathrm{ch}\,x}=\frac{2}{e^x+e^{-x}}$	
csch x 或 cosech x	双曲余割	hyperbolic cosecant	$\mathrm{csch}\,x=\frac{1}{\mathrm{sh}\,x}=\frac{2}{e^x-e^{-x}}$	
arsh x	反双曲正弦	inverse hyperbolic sine	$\mathrm{arsh}\,x=\ln(x+\sqrt{x^2+1})(-\infty<x<+\infty)$	亦可用 arsinh x 表示
arch x	反双曲余弦	inverse hyperbolic cosine	$\mathrm{arch}\,x=\pm\ln(x+\sqrt{x^2-1})\ (x\geqslant 1)$	亦可用 arcosh x 表示
arth x	反双曲正切	inverse hyperbolic tangent	$\mathrm{arth}\,x=\frac{1}{2}\ln\frac{1+x}{1-x}\ (-1<x<1)$	亦可用 artanh x 表示
arcoth x	反双曲余切	inverse hyperbolic cotangent	$\mathrm{arcoth}\,x=\frac{1}{2}\ln\frac{x+1}{x-1}\ (\lvert x\rvert>1)$	
arsech x	反双曲正割	inverse hyperbolic secant	$\mathrm{arsech}\,x=\ln(1\pm\sqrt{1-x^2})-\ln x(0<x\leqslant 1)$	
arcsch x	反双曲余割	inverse hyperbolic cosecant	$\mathrm{arcsch}\,x=\ln(1+\sqrt{1+x^2})-\ln x$	亦可用 arcosech x 表示
$f(x)$	函数	function	如 $y=f(x)$ 表示以 x 为自变量的一元函数	
$f(x_1,\cdots,x_n)$	n 元函数	n-ary function	表示以 $x_1,x_2,\cdots,x_n$ 为自变量的 n 元函数	
Grf	图像	graph	表示函数 f 的图像	
$f(x)\mid_{x=a}$	函数值	function value	表示函数 $f(x)$ 在点 a 处的函数值，即 $f(x)\mid_{x=a}=f(a)$	
$f(x)\mid_a^b$ 或 $[f(x)]_a^b$	函数值的差	difference of the function value	表示函数 $f(x)$ 在区间$[a,b]$端点处函数值的差，即 $f(x)\mid_a^b=f(b)-f(a)$ 或$[f(x)]_a^b=f(b)-f(a)$	这种表示法常用于定积分的计算
const	常值函数	constant function	若 $f(x)=c$，则称 $f(x)$是常值函数，记为 const$f$	亦简记为 $f(x)=c$
$I(x)$	恒等函数	identity function	表示对 D 中一切 x 都有 $I(x)=x$	
$g\circ f$	复合函数	composite function	表示由函数 $f(x)$ 与 $g(x)$ 复合而成的函数，即 $(g\circ f)(x)=g(f(x))$	亦称合成函数
$\rightarrow$	趋于或收敛于	converges to	$x\rightarrow a$ 表示 x 无限接近 a；$x_n\rightarrow a$ 表示序列$\{x_n\}$收敛于 a	$x\not\rightarrow a$ 表示 x 不趋于 a，$x_n\not\rightarrow a$ 表示序列$\{x_n\}$不收敛于 a
$\rightrightarrows$	一致收敛	uniformly convergent	$f_n\rightrightarrows f$ 表示 f_n 在 D 内一致收敛于 f，即 $\lim\limits_{n\to\infty}\sup\limits_{x\in D}\lvert f_n(x)-f(x)\rvert=0$	
$\downarrow$，$\searrow$	单调递减	monotone decreasing	随自变量 x 的增加，函数值 $f(x)$ 逐渐减少	
$\uparrow$，$\nearrow$	单调增加	monotone increasing	随自变量 x 的增加，函数值 $f(x)$ 逐渐增加	
$\simeq$	渐近等于	asymptotically equal to	在某极限过程中，值可以无限接近的两个函数. 如当 $x\rightarrow a$ 时，$\frac{1}{\mathrm{sim}(x-a)}\simeq\frac{1}{x-a}$	在无穷小量比较时，表示等价无穷小，记为 ～
$\lim\limits_{x\to a}f(x)$	极限	limit	$\lim\limits_{x\to a}f(x)=b$ 表示当 x 趋于 a 时，$f(x)$无限接近于 b. 右极限和左极限分别记为：$\lim\limits_{x\to a^+}f(x)$和 $\lim\limits_{x\to a^-}f(x)$	亦可记为：当 $x\rightarrow a$ 时，$f(x)\rightarrow b$
$O(g(x))$	兰道记号	Landau's notation	$f(x)=O(g(x))$意为$\lvert f(x)/g(x)\rvert$在行文所述的极限中有上界	比较无穷小量时，表示同阶无穷小
$o(g(x))$	兰道记号	Landau's notation	$f(x)=o(g(x))$表示在行文所述的极限中 $f(x)/g(x)\rightarrow 0$	比较无穷小量时，表示高阶无穷小
Δx	增量	increment	$\Delta x=x-x_0$ 表示自变量 x 的增量	亦称 x 的改变量

符号	中文名称	英文名称	意义或举例	备注
$\frac{df}{dx}$	导函数或微商	derived function	函数 f 的改变量与自变量 x 的改变量之比，当自变量改变量 Δx 趋于零时的极限表示为 $\lim\limits_{\Delta x\to 0}\frac{\Delta f}{\Delta x}=\frac{df}{dx}$ 或 $\frac{df}{dx}$	亦可用 f' 或 Df 来表示. 简称导数
$\left(\frac{df}{dx}\right)_{x=a}$	导函数值	value of derived function	函数 $f(x)$ 在某点 a 的导数值. 记为 $\left(\frac{dt}{dx}\right)_{x=a}$ 或 $\left(\frac{df'}{dx}\right)_{x=a}$	亦可用 $f'(a)$ 或 $Df(a)$ 来表示
$\frac{d^n f}{dx^n}$	n 阶导数	derivative of n-order	对 $f(x)$ 连续求 n 次一阶导数. 记为 $\frac{d^n f}{dx^n}$ 或 $f^{(n)}$. 当 $n=2,3$ 时，常用 f'',f''' 来代替，称为2阶、3阶导数. 如自变量是时间 t，常用 $f''(t)$ 来代替 $\frac{d^2 f}{dt^2}$	亦可用 $f^{(n)}$ 或 $D^n f$ 来表示
$\frac{\partial f}{\partial x}$ 或 $\partial_x f$	偏导数或偏微商	partial derivative	对多元函数的其中一个自变量 x 求导数，其他变量暂视为常数所得的结果	亦可用 $\left(\frac{\partial f}{\partial x}\right)_{y,\cdots}$ 或 f_x 表示
$\frac{\partial^2 f}{\partial x\partial y}$ 或 f_{xy}	混合偏导数	mixed partial derivative	先对 x 求导，再对 y 求导，即 $\frac{\partial^2 f}{\partial x\partial y}=\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$，	
$\frac{\partial^2 f}{\partial x^2}$ 或 f_{xx}	二阶偏导数	partial derivative of 2-order	对 x 连续求二阶导数，其他变量视为常数	
$\frac{\partial^{n+m} f}{\partial x^n \partial y^m}$	$m+n$ 阶偏微商	partial derivative of $(m+n)$-order	函数 f 先对 x 求 n 次偏微商，再对 y 求 m 次偏微商	
$\frac{\partial(u,v,w)}{\partial(x,y,z)}$	函数行列式	functional determinant	表示 u,v,w 对 x,y,z 的函数行列式，其中 $u(x,y,z),v(x,y,z),w(x,y,z)$ 都是多元函数	亦称雅可比行列式(Jacobian 行列式)
df	全微分	total differential	$df(x_2,x_2,\cdots,x_n)=\frac{\partial f}{\partial x_1}dx_1+\frac{\partial f}{\partial x_2}dx_2+\cdots+\frac{\partial f}{\partial x_n}dx_n$	
$^*\mathbb{R}$ 或 $\mathbb{R}^*$	扩张的实数系	extended real number system	把 $+\infty$ 与 $-\infty$ 加到实数系所得的数系	亦可记为 $[-\infty,+\infty]$
$\{a_n\}$	数列	sequence of number	表示数列 $a_1,a_2,\cdots,a_n\cdots$	
$\sum\limits_{n=1}^{\infty}a_n$	无穷级数	infinite series	无穷数列的各项用加号连结而成的表达式	
$\sum\limits_{m=1}^{\infty}\sum\limits_{n=1}^{\infty}a_{mn}$	叠级数	iterated series	各项均为级数的级数，其中 $\{a_{mn}\}$ 称为二重序列	亦称累级数
$\sum\limits_{m,n=1}^{\infty}a_{mn}$	二重级数	double series	把二重序列的项 a_{mn} 按任意次序排列并用加号连结得到的表达式	
$\prod\limits_{n=1}^{\infty}u_n$	无穷乘积	infinite product	把无穷序列 $u_1,u_2,\cdots,u_n\cdots$ 的各项连乘	
$f(a-0)$	左极限	left limit	$f(a-0)=\lim\limits_{x\to a-0}f(x)$	
$f(a+0)$	右极限	right limit	$f(a+0)=\lim\limits_{x\to a+0}f(x)$	
$f'_-(x)$	左导数	left derivative	$f'_-(x)=\lim\limits_{\Delta x\to 0-}\frac{f(x+\Delta x)-f(x)}{\Delta x}$	
$f'_+(x)$	右导数	right derivative	$f'_+(x)=\lim\limits_{\Delta x\to 0+}\frac{f(x+\Delta x)-f(x)}{\Delta x}$	
$\overline{\int_a^b} f(x)dx$	黎曼上积分	Riemann upper integral	$\overline{\int_a^b} f(x)dx=\sup\limits_{P\in\mathscr{P}}S_P(f)$	
$\underline{\int_a^b} f(x)dx$	黎曼下积分	Riemann lower integral	$\underline{\int_a^b} f(x)dx=\inf\limits_{P\in\mathscr{P}}S_P(f)$	
$\int\limits_{D\subset\mathbb{R}^n}f(x)dx$	n 重积分	n-fold integral	$\int\limits_{D\subset\mathbb{R}^n}f(x)dx=\iiint\cdots\int\limits_D f(x_1,x_2,\cdots,x_n)dx_1dx_2\cdots dx_n$	
Vf	变分	variation	$Vf=f_1(x)-f(x)$	
V 或 Var	变差	variation	$V_a^b f$ 或 $\mathrm{Var}_{[a,b]}f$ 表示函数 f 在 $[a,b]$ 上的全变差，当 $a=b$ 时，定义 $V_a^a f=0$；当 $V_a^b f<\infty$ 时，称 f 为 $[a,b]$ 上的有界变差函数	
δJ	泛函 J 的变分	variation of the functional J	泛函 $J[Y]$ 的一阶变分 $\delta J=\left(\frac{\partial J[Y]}{\partial\varepsilon}\right)_{\varepsilon=0}\cdot\varepsilon$	

符号	中文名称	英文名称	意义或举例	备注
Lip 或 lip	李普希茨条件	Lipschitz condition	$f\in\mathrm{lip}\alpha$ 或 $f\in\mathrm{Lip}\alpha$ 表示函数 f 满足 α 阶李普希茨条件	
Δf	一阶向前差分	forward difference of first-order	$\Delta f(x_i)=f(x_i+h)-f(x_i)$	
$\Delta^2 f$	二阶向前差分	forward difference of second-order	$\Delta^2 f(x_i)=\Delta f(x_i+h)-\Delta f(x_i)$	
$\Delta^n f$	n 阶向前差分	forward difference of n-order	$\Delta^n f(x_i)=\Delta^{n-1}f(x_i+h)-\Delta^{n-1}f(x_i)$	
∇f	一阶向后差分	backward difference of first-order	$\nabla f(x_i)=f(x_i)-f(x_i-h)$	
$\nabla^2 f$	二阶向后差分	backward difference of second-order	$\nabla^2 f(x_i)=\nabla f(x_i)-\nabla f(x_i-h)$	
$\nabla^n f$	n 阶向后差分	backward difference of n-order	$\nabla^n f(x_i)=\nabla^{n-1}f(x_i)-\nabla^{n-1}f(x_i-h)$	
δf	一阶中心差分	centered difference of first-order	$\delta f(x_i)=f\left(x_i+\frac{h}{2}\right)-f\left(x_i-\frac{h}{2}\right)$	
$\delta^2 f$	二阶中心差分	centered difference of second-order	$\delta^2 f(x_i)=\delta f(x_i+\frac{h}{2})-\delta f(x_i-\frac{h}{2})$	
$\delta^n f$	n 阶中心差分	centered difference of n-order	$\delta^n f(x_i)=\delta^{n-1}f\left(x_i+\frac{h}{2}\right)-\delta^{n-1}f\left(x_i-\frac{h}{2}\right)$	
$\int f(x)\mathrm{d}x$	不定积分	indefinite integral	$\int f(x)\mathrm{d}x=F(x)+C$，其中 $F(x)$ 是 $f(x)$ 在 $[a,b]$ 上的一个原函数，C 是任意常数	
$\int_a^b f(x)\mathrm{d}x$	定积分	definite integral	$\int_a^b f(x)\mathrm{d}x=\lim\limits_{\lambda\to 0}\sum\limits_{i=1}^{n}f(\xi_i)\Delta x_i$，其中 $\lambda=\max\limits_i\{\Delta x_i\}$	
P.V. $\int_a^b f(x)\mathrm{d}x$	柯西主值	Cauchy principal value	$P.V.\int_a^b f(x)\mathrm{d}x$ $=\lim\limits_{\varepsilon\to 0^+}\left(\int_a^{c-\varepsilon}f(x)\mathrm{d}x+\int_{c+\varepsilon}^b f(x)\mathrm{d}x\right)$ 或 $P.V.\int_a^b f(x)\mathrm{d}x=\lim\limits_{M\to\infty}\int_{-M}^{M}f(x)\mathrm{d}x$	
$\int_C,\int_S,\int_V,\oint$	积分号	sign of integration	$\int_C,\int_S,\int_V,\oint$ 分别表示沿曲线 C，沿曲面 S，沿体积 V 以及沿闭曲线或闭曲面的积分	
$C(z),S(z)$	菲涅耳积分	Fresnel integral	$C(z)=\frac{1}{\sqrt{2\pi}}\int_0^z\frac{\cos t}{\sqrt{t}}\mathrm{d}t$, $S(z)=\frac{1}{\sqrt{2\pi}}\int_0^z\frac{\sin t}{\sqrt{t}}\mathrm{d}t$	
$\iint\limits_D f(x,y)\mathrm{d}x\mathrm{d}y$	二重积分	double integral	二元函数 $f(x,y)$ 在平面区域 D 上的积分	
$\mathrm{Li}(x)$ 或 $\mathrm{li}(x)$	对数积分	logarithmic integral	$\mathrm{Li}(x)=\int_0^x\frac{\mathrm{d}t}{\log t}$，高斯用函数 $\frac{1}{\log t}$ 表示在大整数 t 附近的素数分布的平均密度	
$\mathrm{Ei}(x)$	指数积分	exponential integral	$\mathrm{Ei}(x)=\int_x^{\infty}\frac{\mathrm{e}^{-t}}{t}\mathrm{d}t$，当 $x<0$ 时，在 $t=0$ 处取积分主值	在量子力学中有重要应用
$\mathrm{Si}(x)$	正弦积分	sine integral	$\mathrm{Si}(x)=\int_0^x\frac{\sin t}{t}\mathrm{d}t$	在通信工程中有重要应用
$\mathrm{Ci}(x)$	余弦积分	cosine integral	$\mathrm{Ci}(x)=-\int_x^{\infty}\frac{\cos t}{t}\mathrm{d}t$	在通信工程中有重要应用
$\mathrm{sgn}\,x$	符号函数	sign function	当 $x\in\mathbf{R}$ 时，$\mathrm{sgn}\,x=\begin{cases}1 & (x>0),\\ 0 & (x=0),\\ -1 & (x<0);\end{cases}$ 当 $x\in\mathbf{C}$ 时，$\mathrm{sgn}\,x=\begin{cases}\frac{x}{\lvert x\rvert} & (x\neq 0),\\ 0 & (x=0)\end{cases}$	亦称克罗内克函数
ε_{ijk}	列维-齐维塔符号	Levi-Civita symbol	$\varepsilon_{ijk}=\begin{cases}1 & (\text{若 } ijk \text{ 为 } 1,2,3 \text{ 的偶排列}),\\ -1 & (\text{若 } ijk \text{ 为 } 1,2,3 \text{ 的奇排列}),\\ 0 & (\text{若 } ijk \text{ 为 } 1,2,3 \text{ 的真重复排列})\end{cases}$	
$\varepsilon(x)$	单位阶跃函数或称赫维赛德函数	unit step function or Heaviside function	$\varepsilon(x)=\begin{cases}1 & (x\geqslant 0),\\ 0 & (x<0)\end{cases}$ 视作广义函数时的定义为 $\varepsilon(x)=\begin{cases}1 & (x>0),\\ 0 & (x<0)\end{cases}$	亦可用 $H(x)$ 表示

符号	中文名称	英文名称	意义或举例	备注
$f*g$	f 与 g 的卷积	convolution of f and g	$(f*g)(x)=\int_{-\infty}^{\infty}f(y)g(x-y)\mathrm{d}y$，式中 $f(x)$ 和 $g(x)$ 是 $(-\infty,\infty)$ 内的绝对可积函数	
sn x cn x dn x	雅可比椭圆函数	Jacobi elliptic function	$\mathrm{sn}\,x=\sqrt{e_1-e_3}\dfrac{\sigma(u)}{\sigma_3(u)}$；$\mathrm{cn}\,x=\dfrac{\sigma_1(u)}{\sigma_3(u)}$； $\mathrm{dn}\,x=\dfrac{\sigma_2(u)}{\sigma_3(u)}$，其中 $x=u\sqrt{e_1-e_3}$	
$\wp(x)$	外尔斯特拉斯椭圆函数	Weierstrass's elliptic function	$\wp(x)=\dfrac{1}{x^2}+\sum\limits_{\omega\neq0}\left(\dfrac{1}{(x-\omega)^2}-\dfrac{1}{\omega^2}\right)$	
B_n 或 b_n	伯努利数	Bernoulli's numbers	解析函数 $(e^z-1)^{-1}$ 在 $z=0$ 附近的罗朗级数展开式 $\dfrac{1}{z}-\dfrac{1}{2}+\sum\limits_{n=1}^{\infty}(-1)^{n-1}\dfrac{B_n}{(2n)!}z^{2n-1}$， 则称式中系数 B_n 为伯努利数	
supp f 或 spt f	函数的支集	support of function	若 Ω 是局部紧空间，则 Ω 上函数 f 的支集是 Ω 中的集合 $\{x\mid f(x)\neq0\}$ 的闭包，表示成 supp$f$	
$\delta(x)$	狄拉克函数	Dirac δ-function	质量分布在区域 Ω 的总量为 $\iiint\limits_{\Omega}\delta_{M_0}(M)\mathrm{d}M=\begin{cases}1 & (M_0\in\Omega),\\0 & (M_0\overline{\in}\Omega),\end{cases}$ 称这样的函数为 $\delta(x)$ 函数，它在每一点的值 $\delta_{M_0}(M)=\begin{cases}0 & (M_0\neq M),\\\infty & (M_0=M)\end{cases}$	亦称 δ 函数
am x	振幅函数	amplitude function	在形如 $I_\varphi(au)=\iint e^{i\varphi(x,\theta)}a(x,\theta)u(x)\mathrm{d}x\mathrm{d}\theta$ 的振荡积分中，$a(x,\theta)$ 称为振幅函数	
$\Gamma(x)$	伽马函数	gamma function	$\Gamma(x)=\int_0^{\infty}t^{x-1}e^{-t}\mathrm{d}t\ (x>0)$， $\Gamma(n+1)=n!\ (n=0,1,2,\cdots)$	亦称 Γ 函数
$\gamma(x)$	不完全伽马函数	incomplete gamma function	$\gamma(x)=\int_0^{\lambda}e^{-t}t^{x-1}\mathrm{d}t$；$\gamma(x)=\int_{\lambda}^{\infty}e^{-t}t^{x-1}\mathrm{d}t$，其中 x>0	在统计学和分子结构论中常用
$B(x,y)$	贝塔函数	beta function	$B(x,y)=\int_0^1t^{x-1}(1-t)^{y-1}\mathrm{d}t$，$(x,y\in\mathrm{R};x>0,y>0)$；$B(x,y)=\dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$	亦称 β 函数
$\Psi(x)$	普西函数	psi function	$\Psi(x)=\dfrac{\mathrm{d}}{\mathrm{d}x}(\ln\Gamma(x))$ 是函数方程 $\Psi(x+1)-\Psi(x)=\dfrac{1}{x}$，$\Psi(1)=-c$，$\lim\limits_{n\to\infty}(\Psi(x+n)-\Psi(1+n))=0$ 的解	亦称 Ψ 函数
$F(k,\varphi)$	第一类不完全椭圆积分	incomplete elliptic integral of the first kind	$F(k,\varphi)=\int_0^{\varphi}\dfrac{\mathrm{d}\varphi}{\sqrt{1-k^2\sin^2\varphi}}$	
$E(k,\varphi)$	第二类不完全椭圆积分	incomplete elliptic integral of the second kind	$E(k,\varphi)=\int_0^{\varphi}\sqrt{1-k^2\sin^2\varphi}\mathrm{d}\varphi$	
$\Pi(n,k,\varphi)$	第三类不完全椭圆积分	incomplete elliptic integral of the third kind	$\Pi(n,k,\varphi)=\int_0^{\varphi}\dfrac{\mathrm{d}\varphi}{(1+n\sin^2\varphi)\sqrt{1-k^2\sin^2\varphi}}$	
$K(k)$	第一类完全椭圆积分	complete elliptic integral of the first kind	$K(k)=F(k,\pi/2)=\int_0^{\pi/2}\dfrac{\mathrm{d}\varphi}{\sqrt{1-k^2\sin^2\varphi}}$	
$E(k)$	第二类完全椭圆积分	complete elliptic integral of the second kind	$E(k)=E(k,\pi/2)=\int_0^{\pi/2}\sqrt{1-k^2\sin^2\varphi}\mathrm{d}\varphi$	
$\Pi(n,k,\pi/2)$	第三类完全椭圆积分	complete elliptic inlegral of the third kind	$\Pi(n,k,\pi/2)=\int_0^{\pi/2}\dfrac{\mathrm{d}\varphi}{(1+n\sin^2\varphi)\sqrt{1-k^2\sin^2\varphi}}$	
$P_l(x)$	勒让德多项式	Legendre polynomial	方程 $(1-x^2)y''-2xy'+l(l+1)y=0$ 的特解， $P_l(x)=\sum\limits_{r=0}^{[\frac{n}{2}]}(-1)^r\dfrac{(2n-2r)!}{2^nr!(n-r)!(n-2r)!}x^{n-2r}$	

符号	中文名称	英文名称	意义或举例	备注
$P_l^m(x)$	关联勒让德函数	associated Legendre function	方程 $(1-x^2)y''-2xy'+[l(l+1)-\frac{m^2}{1-x^2}]y=0$ 的特解，$P_l^m(x)=(-1)^m(1-x^2)^{\frac{m}{2}}\frac{\mathrm{d}^m}{\mathrm{d}x^m}P_l(x)\ (l,m=0,1,2,\cdots;m\leqslant l)$	
$T_n(x)$	第一类切比雪夫多项式	Chebyshev polynomial of the 1st kind	方程 $(1-x^2)y''-xy'+n^2y=0$ 的特解，$T_n(x)=\cos(n\arccos x)\ (n=0,1,2,\cdots)$	
$U_n(x)$	第二类切比雪夫多项式	Chebyshev polynomial of the 2nd kind	方程 $(1-x^2)y''-3xy'+n(n+2)y=0$ 的特解，$U_n(x)=\frac{\sin[(n+1)\arccos x]}{\sin(\arccos x)}(n=0,1,2,\cdots)$	
$L_n(x)$	拉盖尔多项式	Laguerre polynomial	方程 $xy''+(1-x)y'+ny=0$ 的特解，$L_n(x)=\frac{\mathrm{e}^x}{n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}(x^n\mathrm{e}^{-x})\ (n=0,1,2,\cdots)$	
$H_n(x)$	埃尔米特多项式	Hermite polynomial	方程 $y''-2xy'+2ny=0$ 的特解，$H_n(x)=(-1)^n\mathrm{e}^{x^2}\frac{\mathrm{d}^n}{\mathrm{d}x^n}\mathrm{e}^{-x^2}(n=0,1,2,\cdots)$	
H_c	超平面	hyperplane	$H_c=\{x\in\mathrm{R}^n:\langle a,x\rangle=c\}$，式中 c 为实数，a 为 R^n 中的非零元	
$F(a;b;c;x)$	超几何函数	hypergeometric function	方程 $x(1-x)y''+[c-(a+b+1)x]y'-aby=0$ 的特解，$F(a;b;c;x)=1+\frac{ab}{c}x+\frac{a(a+1)b(b+1)}{2!c(c+1)}x^2+\cdots$	亦称超比函数
$F(a;c;x)$	合流超几何函数	hypergeometric function of confluent type	方程 $xy''+(c-x)y'-ay=0$ 的特解，$F(a;c;x)=1+\frac{a}{c}x+\frac{a(a+1)}{2!c(c+1)}x^2+\cdots$	亦称汇合型超几何函数或库默尔函数
$J_l(x)$	第一类柱贝塞尔函数	cylindrical Bessel function of the 1st kind	方程 $x^2y''+xy'+(x^2-l^2)y=0$ 的特解，$J_l(x)=\sum_{k=0}^{\infty}\frac{(-1)^k(x/2)^{l+2k}}{k!\Gamma(l+k+1)}$	
$N_l(x)$	第二类柱贝塞尔函数	cylindrical Bessel function of the 2nd kind	$N_l(x)=\lim_{k\to l}\frac{J_k(x)\cos k\pi-J_{-k}(x)}{\sin k\pi}$. 它是贝塞尔方程的第二解，可由第一类柱贝塞尔函数定义	亦称柱汉克尔函数
$H_l^{(1)}(x)$ $H_l^{(2)}(x)$	第三类柱贝塞尔函数	cylindrical Bessel function of the 3rd kind or cylindrical Hankel function	$H_l^{(1)}(x)=J_l(x)+\mathrm{i}N_l(x)$，$H_l^{(2)}(x)=J_l(x)-\mathrm{i}N_l(x)$. 它们是第一类和第二类柱贝塞尔的线性组合，是贝塞尔方程的两个线性无关解	亦称柱汉克尔函数
$I_l(x)$ $K_l(x)$	修正的柱贝塞尔函数	modified cylindrical Bessel function	方程 $x^2y''+xy'-(x^2+l^2)y=0$ 的特解，$I_l(x)=\mathrm{i}^{-l}J_l(\mathrm{i}x)$，$K_l(x)=\left(\frac{\pi}{2}\right)\mathrm{i}^{l+1}[J_l(\mathrm{i}x)+\mathrm{i}N_l(\mathrm{i}x)]$	亦称变形的柱贝塞尔函数
$j_l(x)$	第一类球贝塞尔函数	spherical Bessel function of the 1st kind	方程 $x^2y''+2xy'+[x^2-l(l+1)]y=0$ 的特解，$j_l(x)=\left(\frac{\pi}{2x}\right)^{\frac{1}{2}}J_{l+\frac{1}{2}}(x)$	
$n_l(x)$	第二类球贝塞尔函数	spherical Bessel function of the 2nd kind	$n_l(x)=\left(\frac{\pi}{2x}\right)^{\frac{1}{2}}N_{l+\frac{1}{2}}(x)$	亦称球诺伊曼函数，也记为 $y_l(x)$
$h_l^{(1)}(x)$ $h_l^{(2)}(x)$	第三类球贝塞尔函数	spherical Bessel function of the 3rd kind	$h_l^{(1)}(x)=j_l(x)+\mathrm{i}n_l(x)=\left(\frac{\pi}{2x}\right)^{\frac{1}{2}}H_{l+\frac{1}{2}}^{(1)}(x)$，$h_l^{(2)}(x)=j_l(x)-\mathrm{i}n_l(x)=\left(\frac{\pi}{2x}\right)^{\frac{1}{2}}H_{l+\frac{1}{2}}^{(2)}(x)$	修正的球贝塞尔函数，分别记为 $i_l(x)$ 与 $k_l(x)$
∇	矢量微分算子	operator of vector differentiation	$\nabla=\boldsymbol{e}_x\frac{\partial}{\partial x}+\boldsymbol{e}_y\frac{\partial}{\partial y}+\boldsymbol{e}_z\frac{\partial}{\partial z}=\boldsymbol{e}_i\frac{\partial}{\partial x_i}$	亦称哈密顿算子
grad，∇	梯度	gradient	若 $f:D(\subseteq\mathrm{R}^n)\to\mathrm{R}$，则 f 在 $\boldsymbol{a}\in D$ 的梯度为 $\mathrm{grad}f(\boldsymbol{a})=\left(\frac{\partial f}{\partial x_1}(\boldsymbol{a}),\frac{\partial f}{\partial x_2}(\boldsymbol{a}),\cdots,\frac{\partial f}{\partial x_n}(\boldsymbol{a})\right)$	
div，$\nabla\cdot$	散度	divergence	若向量函数 $\boldsymbol{f}(x,y,z)=(P,Q,R)$ 连续可微，则向量场的散度为 $\mathrm{div}\boldsymbol{f}=\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}$	
rot，$\nabla\times$	旋度	rotation	$\boldsymbol{f}=(P,Q,R)$ 是三维向量函数，$\boldsymbol{f}$ 的旋度为 $\mathrm{rot}\boldsymbol{f}=\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z},\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x},\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)$	

数学符号表

符号	中文名称	英文名称	意义或举例	备注
Δ, ∇^2	拉普拉斯算子	Laplacian operator	$\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$	亦称调和算子
$\square$	达朗贝尔算子	d'Alembertain operator	$\square=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}$	c 为电磁波在真空中的传播速度
D	微分算子	differential operator	即 $\frac{\mathrm{d}f(t)}{\mathrm{d}t}=Df(t)$，$\frac{\mathrm{d}^2f(t)}{\mathrm{d}t^2}=D^2f(t)$，…，$\frac{\mathrm{d}^nf(t)}{\mathrm{d}t^n}=D^nf(t)$	
Λ	拓扑双曲不变集	topological hyperbolic set	$f:M\to M$ 是微分同胚．f 的不变闭子集 $\Lambda\subset M$ 称为拓扑双曲不变集	
Diff′	微分同胚空间	differential homeomorphic space	Diff′(M)表示 M 全体微分同胚构成的空间	
Homeo	同胚空间	homeomorphic space	Homeo(M)表示 M 的全体同胚构成的空间	
Proj	射影基向量	base vector of projective	Proj$k$ 表示 P—标架的第 k 个基向量	
Ob	阻碍集	obstruction sets	Ob(S)表示向量场 S 的阻碍集	
Log$z$	对数函数	logarithmic function	$w=\mathrm{Log}z=\log\lvert z\rvert+\mathrm{i}(\arg z+2k\pi)\ (k=0,\pm1,\pm2,\cdots)$，$z$ 为复数	
sin z	复变正弦函数	sine function of a complex variable	$\sin z=\frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}z}-\mathrm{e}^{-\mathrm{i}z})$，式中 z 为复变数．当 z 为实数时与数学分析中的正弦函数的定义一致	
cos z	复变余弦函数	cosine function of a complex variable	$\cos z=\frac{1}{2}(\mathrm{e}^{\mathrm{i}z}+\mathrm{e}^{-\mathrm{i}z})$，式中 z 为复变数．当 z 为实数时与数学分析中的余弦函数的定义一致	
tan z	复变正切函数	tangent function of a complex variable	$\tan z=\frac{\sin z}{\cos z}$	
Arc sin z	复变反正弦函数	inverse sine function of a complex variable	$\mathrm{Arc}\sin z=-\mathrm{i}\log(\mathrm{i}z+\sqrt{1-z^2})$，式中 z 为复变数，$\mathrm{e}^{\mathrm{i}w}=\mathrm{i}z+\sqrt{1-z^2}$	
Arc cos z	复变反余弦函数	inverse cosine function of a complex variable	$\mathrm{Arc}\cos z=-\mathrm{i}\log(z+\mathrm{i}\sqrt{1-z^2})$，式中 z 为复变数	
Arc tan$z$	复变反正切函数	inverse tangent function of a complex variable	$\mathrm{Arc}\tan z=\frac{1}{2i}\log\frac{i-z}{i+z}$，式中 z 为复变数	
L(z)	分式线性变换	fractional linear transformation	$\mathrm{L}(z)=\frac{az+b}{cz+d}$，式中 a,b,c,d 都是复常数，且 $ad-bc\neq0$	若 a,b,c,d 都是实数，且 $ad-bc>0$ 称此为富克斯变换
(a,b,c,d)	交比	cross ratio	$(a,b,c,d)=\frac{c-a}{c-b}:\frac{d-a}{d-b}$，式中 a,b,c,d 是任意四个互异的复数	亦称非调和比
$n(\gamma;a)$	环绕数	winding number	点 a 关于 γ 的环绕数，$n(\gamma;a)=\frac{1}{2\pi\mathrm{i}}\int_\gamma\frac{\mathrm{d}\zeta}{\zeta-a}$，式中 γ 是一条可求长的闭路径，a 点不在 γ 上	亦称指示数或卷绕数
Res$f(z)$	留数	residue	在 $f(z)$ 的孤立奇点 a 的去心邻域内的罗朗级数展开式中，$1/(z-a)$ 项的系数为 c_{-1}，即 $\underset{z=a}{\mathrm{Res}}f(z)=c_{-1}=\frac{1}{2\pi\mathrm{i}}\int_{\substack{\lvert z-a\rvert=\rho\\(0<\rho<R)}}f(z)\mathrm{d}z$	亦称残数
$L(s)$	拉普拉斯变换	Laplace transform	$f(t)$ 的拉普拉斯变换为 $L(s)=\int_0^\infty f(t)\mathrm{e}^{-st}\mathrm{d}t$	
$F(\xi)$	傅里叶变换	Fourier transform	$f(x)$ 的傅里叶变换为 $F(\xi)=\frac{1}{\sqrt{2\pi}}\int_0^\infty f(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x$	
$F_c(\xi)$	傅里叶余弦变换	Fourier cosine transform	$f(x)$ 的傅里叶余弦变换为 $F_c(\xi)=\sqrt{\frac{2}{\pi}}\int_0^\infty f(x)\cos\xi x\mathrm{d}x$	
$F_s(\xi)$	傅里叶正弦变换	Fourier sine transform	$f(x)$ 的傅里叶正弦变换为 $F_s(\xi)=\sqrt{\frac{2}{\pi}}\int_0^\infty f(x)\sin\xi x\mathrm{d}x$	
$M(z)$	梅林变换	Mellin transform	$f(x)$ 的梅林变换为 $M(z)=\int_0^\infty f(x)x^{z-1}\mathrm{d}x$	

符号	中文名称	英文名称	意义或举例	备注
$H(\xi)$	汉克尔变换	Hankel transform	$f(x)$ 的 v 阶汉克尔变换为 $$H(\xi)=\int_0^{\infty}xf(x)J_v(\xi x)\mathrm{d}x$$	
$G(n)$	勒让德变换	Legendre transform	$f(x)$ 的勒让德变换为 $G(n)=\int_{-1}^{1}f(x)P_n(x)\mathrm{d}x$	
$\mathrm{erf}(z)$	概率积分	probability integral	$\mathrm{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z \mathrm{e}^{-u^2}\mathrm{d}u$	
$\mathrm{erfc}(z)$	余概率积分	complement probability in- tegral	$\mathrm{erfc}(z)=\frac{2}{\sqrt{\pi}}\int_z^{+\infty}\mathrm{e}^{-u^2}\mathrm{d}u$	
$\Phi c(z)$	正态概率积分	normal probability inte-gral	$\Phi_c(z)=\int_{-\infty}^{z}\frac{2}{\sqrt{\pi}}\mathrm{e}^{-\frac{2}{u}}\mathrm{d}u$	
${}_pF_q$	超几何级数	hypergeometric series	超几何级数的一般形式是 $${}_pF_q(\alpha_1,\cdots,\alpha_p,\beta_1,\cdots,\beta_q;z)=\sum_{n=0}^{\infty}\frac{(\alpha_1)_n\cdots(\alpha_p)_nz^n}{(\beta_1)_n\cdots(\beta_q)_nn!}$$	
E_n 或 γ	欧拉常数	Euler constant	$\gamma=\lim\limits_{n\to\infty}\left(1+\frac{1}{2}+\cdots+\frac{1}{n}-\ln n\right)$ $\approx 0.5772156649015328606065 1209\cdots$	
$\{f,D\}$	解析函数元素	holomorphic function el-ement	复平面上的区域 D 连同在其内全纯的一个函数 $f(z)$，合成为解析函数元素	简称函数元素
$k(z)$	克贝函数	Koebe function	$k(z)=z(1-z)^{-2}$，$k_\theta(z)=\mathrm{e}^{i\theta}k(\mathrm{e}^{i\theta}z)$，$\beta=\lim\limits_{n\to\infty}\|a_n\|/n\leqslant 1$. 其中 $k(z)$ 是 S 类上许多泛函极值问题的极值函数，称 $k_\theta(z)$ 为克贝函数的旋转	
$I_p(r)$	哈代凸性函数	Hardy's convexity func-tion	$I_p(r)=\frac{1}{2\pi}\int_0^{2\pi}\|f(re^{i\theta})\|^p\mathrm{d}\theta\quad(0<r<R)$	
B	布洛赫常数	Bloch's constant	$B=\inf\{\beta(f)\|f\in\mathscr{F}\}$，式中 $\beta(f)=\sup\{r\|r$ 是 $f(\triangle)$ 所包含的单叶圆的半径$\}$	已经证明 $\sqrt{3}/4\leqslant B\leqslant 0.47$
L	兰道常数	Landau's constant	$L=\inf\{\lambda(f),f\in\mathscr{F}\}$，式中 $\lambda(f)=\sup\{r\|r$ 是 $f(\triangle)$ 所包含圆的半径，$f\in\mathscr{F}\}$	已经证明 $0.5\leqslant L\leqslant 0.54326$
$M(\Gamma)$	曲线族 Γ 的模	module of a family of curves Γ	$M(\Gamma)=\inf\limits_{\rho\in p(\Gamma)}\int_D\rho^2\|\mathrm{d}z\|$，其中 Γ 是平面区域 D 上的若尔当曲线族，ρ 是定义在 D 上的非负波莱尔函数	
$\mathrm{M}(f(\Gamma))$	拟共形映射	quasiconformal mapping	f 满足 Beltrami 微分方程 $f_{\bar{z}}=\mu f_z$，称 f 为 μ 共形映射，如 $\\|\mu\\|_\infty<1$，则称 f 为拟共形映射	亦称拟保角映射
$w(z,a,D)$	调和测度	harmonic measure	a 关于区域 D 的调和测度 $w(z,a,D)$ 是 z 对 (a,b) 的视角. $w(z,a,D)=\frac{1}{\pi}\arg\frac{b-z}{a-z}$	$0\leqslant w(z,x,d)\leqslant 1$
$g(z,a)$ 或 $\mathrm{G}(z,a)$	格林函数	Green's function	函数 $g(z,a)$ 在 D 内奇点 a 的格林函数 $$g(z,a)=\log\left\|\frac{z-\bar{a}}{z-a}\right\|$$	
$\mathrm{E}(z,p)$	外尔斯特拉斯基本因式	Weierstrass basis factor	$E(z,p)=(1-z)\exp\left\{z+\frac{z^2}{2}+\cdots+\frac{z^p}{p}\right\}$	
$T(r,f)$	奈望林纳特征函数	Nevanlinna's character-istic function	满足 $\overline{\lim\limits_{r\to\infty}}\frac{T(r,f)}{(\log r)^2}=\infty$ 的 $T(r,f)$ 是 $f(z)$ 的奈望林纳特征函数	亦称奈望林纳记号，可记为 $T(r)$
$n(r,a)$	a 点个数	number of a-point	$n(r,a)$ 是方程 $f(x)=a$ 在 $\|z\|\leqslant r$ 内解的个数(包括计算重数)	
$\delta(a)$	亏量	defect	$w(z)$ 关于 a 的亏量 $\delta(a)=1-\overline{\lim\limits_{r\to\infty}}\frac{n(r,a)}{T(r,w)}$	亦称亏值
$\mathring{\mathrm{T}}(r,w)$	球面特征函数	spherical characteristic function	$\mathring{T}(r,w)=\frac{1}{\nu}\int_0^r\frac{A(t,w)}{t}\mathrm{d}t$，式中 $w(z)$ 为代数体函数	
$M(r,f)$	整函数的最大模	maximum modulus of entire function	$f(z)$ 的最大模 $M(r,f)=\max\limits_{\|z\|\leqslant r}\|f(z)\|$；$f(z)$ 的 p 次整函数的模 $$M_p(r,f)=\left(\frac{1}{2\pi}\int_0^{2\pi}\|f(re^{i\theta})\|d\theta\right)^{1/p}(0<p<+\infty);$$ 超越整函数 $f(z)$ 的最大模 $M_\infty(r,f)=\max\limits_{\|z\|=r}\|f(z)\|_{p=+\infty}$	

符号	中文名称	英文名称	意义或举例	备注
$B(z)$	布拉施克乘积	Blaschke product	$B(z)=\prod_{n=1}^{\infty}\frac{\lvert a_n\rvert}{a_n}\left(\frac{a_n-a}{1-\bar{a}_n z}\right)$，式中$a_n(n=1,2,\cdots)$是复数序列，$0<a_n<1$	
H^p	哈代空间	Hardy space	所有哈代函数构成的空间，即$H^p(D)=\{f\mid f(z)$在D内解析，$\sup_{0\leqslant r<1}\left(\int_0^{2\pi}\lvert f(re^{i\theta})\rvert^p d\theta\right)^{1/p}<+\infty\}$，其中$D=\{z\mid \lvert z\rvert<1\}$	H^p是由哈代于1915年提出的
$S(z)$	奇异内函数	singular inner function	$S(z)=\exp\{-\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{it}+z}{e^{it}-z}d\mu(t)\}$，式中$\mu(t)$是非减的有界变差函数，其导数几乎处处等于零	
$F(z)$	外函数	outer function	$F(z)=\exp\left\{\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{it}+z}{e^{it}-z}\log\lvert f(e^{it})\rvert dt\right\}$	
BMOA	有界平均振荡解析函数类	analysis function class of the bounded mean oscillating	$\mathrm{BMOA}(D)=\{f\mid f(z)$是单位圆周$T$上的可积函数，$u(e^{i\theta})$的积分$\sup_{T\subset I}\frac{1}{\lvert I\rvert}\int_I f(u-u_I)d\theta<+\infty\}$，式中$u$为单圆周$T$上的可积函数，$I$是$T$的子弧，$\lvert I\rvert$是$I$的长度	
B_n	C^n中单位球	unit ball in a C^n	$B_n=\{z=(z_1,z_2,\cdots,z_n)\mid \lvert z_1\rvert^2+\lvert z_2\rvert^2+\cdots+\lvert z_n\rvert^2<1\}$	
$\mathrm{Aut}(D)$	域的全纯自同构群	holomorphic automorphism group of a domain	表示域D的全纯自同构的全体组成的群．它是D上的拓扑变换群	
∂D	域的边界	boundary of a domain	域D和它的闭包$\bar{D}$的差集，即$\partial D=\bar{D}\backslash \dot{D}$	
$\mathrm{Hol}(D)$	全纯复线性空间	holomorphic complex linear space	表示D上所有全纯函数构成的复线性空间	
$\bar{\partial}$	$\bar{\partial}$算子	$\bar{\partial}$-operator	$\bar{\partial}:C^1(D)\to L_{\mathfrak{H}}^2,u\mapsto\left(\frac{\partial u}{\partial x_1},\frac{\partial u}{\partial x_2},\cdots,\frac{\partial u}{\partial x_n}\right)$称为$\bar{\partial}$算子	
$H(z,\bar{z})$	正定埃尔米特方阵	positive definite Hermitian matrix	$H(z,\bar{z})=\begin{pmatrix}h_{11}(z,\bar{z}) & \cdots & h_{1n}(z,\bar{z})\\ \vdots & & \vdots\\ h_{n1}(z,\bar{z}) & \cdots & h_{nn}(z,\bar{z})\end{pmatrix}$，式中$h_{jk}(z,\bar{z})$在拓扑积$\varphi_\alpha(U_x)\times\overline{\varphi_\alpha(U_\alpha)}$上全纯	互逆正定埃尔米特方阵记为$\tilde{H}(z,\bar{z})$
$B_\mu^2(M)$	可测复线性空间	measurable complex linear space	$B_\mu^2(M)=\mathrm{Hol}(M)\cap L_\mu^2(M)$，其中M为n维复流形，$\mu$为$M$上任给的测度	
$N(\Omega)$	奈望林纳函数类	Nevanlinna function class	Ω是C^n中的对称域，b是特征边界，若$\Omega\to Cf$在Ω中全纯，且满足$\sup_{0<r<1}\int_b\log^+\lvert f(r,\zeta)\rvert d\sigma(\zeta)<+\infty$，则$f$属于奈望林纳函数类	
$\beta(\Omega)$	布洛赫空间	Bloch space	Ω上全体布洛赫函数的集合，称为布洛赫空间．Ω是C^n中齐线性有界域	
$\rho(\ ,\ )$	点集的距离	distance between two point sets	$\rho(A,B)=\inf_{\substack{x\in A\\ y\in B}}\{\rho(x,y)\}$	
F_σ	F_σ型集	set of type F_σ	表示可数个闭集的并集	F_σ是波莱尔集
G_δ	G_δ型集	set of type G_δ	表示可数个开集的交集	G_δ是波莱尔集
mE；$\lvert E\rvert$	勒贝格测度	Lebesgue measure	若$E\subset R^n$为勒贝格可测集，则E的勒贝格外测度称为勒贝格测度	
$m^*(E)$；$\lvert E\rvert_e$	勒贝格外测度	Lebesgue outer measure	$m^*(E)=\inf\{\sum_{i\in N}\lvert I_i\rvert\mid\{I_i\}$为覆盖$E$的可数个开集$\}$	
$m_*(E)$；$\lvert E\rvert_i$	勒贝格内测度	Lebesgue inner measure	$m_*(E)=\sup\{m(F)\mid F$为闭集，且$F\subset E\}$	
$\aleph_0$	可列集的势	cardinal number of countable set	每一个无穷集的势都是某个阿列夫，自然数集的势是$\aleph_0$	
$\aleph$或C	连续集的势	cardinal number of continuous set	与区间[0,1]对等的集的势记为N或C．连续集的势$C=2^{\aleph_0}$	亦称基数
CH	连续统假设	continuum hypothesis	康托尔猜测：实数集的一切无穷子集或者与自然数集等势或者与连续统等势	

符号	中文名称	英文名称	意义或举例	备注
GCH	广义连续统假设	generalized continuum hypothesis	假设：1. 对任一序数 α，$2^{\aleph_\alpha}=\aleph_{\alpha+1}$； 2. 对任二无穷势 κ,λ，若 $\kappa\leqslant\lambda\leqslant 2^{\aleph}$，则 $\lambda=\kappa$ 或者 $\lambda=2^{\aleph}$	
$H_\alpha(E)$	豪斯多夫测度	Hausdorff measure	$H_\alpha(E)=\lim\limits_{\varepsilon\to 0}H_{\alpha,\varepsilon}(E)=\sup\limits_{\varepsilon>0}H_{\alpha,\varepsilon}(E)$，其中，$H_{\alpha,\varepsilon}(E)=\inf\sum\limits_k\delta(E_k)^\alpha$，且 $\delta(E_k)$ 为 R^n 的子集 E_k 的直径	
$\psi(x)$	狄利克雷函数	Dirichlet function	$\psi(x)=\begin{cases}1, & x\text{ 为有理点},\\ 0, & x\text{ 为无理点}\end{cases}$	亦可用 $D(x)$ 表示
$\chi(n)$或 $\chi_q(n)$或 $\chi(n)\bmod q$	狄利克雷特征	Dirichlet character	整数集上的函数 $\chi(n)=\begin{cases}\exp\left[2\pi\left(\frac{mr}{c}+\frac{m_0r_0}{c_0}+\frac{m_1r_1}{c_1}+\cdots+\frac{m_sr_s}{c_s}\right)\right] & ((n,q)=1)\\ 0 & ((n,p)>1)\end{cases}$	亦称 q 的特征
$\{A,B\}$	泊松符号	Poisson symbol	$\{A,B\}=\sum\limits_{j=1}^{n}\left(\frac{\partial A}{\partial\xi_j}\frac{\partial B}{\partial x_j}-\frac{\partial B}{\partial\xi_j}\frac{\partial A}{\partial x_j}\right)$	亦称泊松括号
$\lvert I\rvert$	I 区间的体积	volume of I-interval	E 为 R^n 中的有界点集，I 为包含 E 的任何有界区间，则以 $\lvert I\rvert$ 表示区间 I 的体积	
a. e. p. p.	几乎处处	almost everywhere	若命题 $P(x)$ 与集合 $E\subset\mathrm{R}^n$ 有关，且零集 $E_0\subset E$，对于任意 $x\in E\backslash E_0$，$P(x)$ 均成立，则称 $P(x)$ 在 E 上几乎处处成立，记为 $P(x)$a. e. 或 $P(x)$p. p.	a. e. 是英文 almost everywhere 的首字母；p. p. 是法文 presque partout 的首字母
$M(x)$	上极限函数	upper limit function	$M(x)=\lim\limits_{\delta\to 0}M(x,\delta)$，其中 $M(x,\delta)$ 为函数 $f(x)$ 在点 x 的 δ 邻域上取值的上确界	
$m(x)$	下极限函数	lower limit function	$m(x)=\lim\limits_{\delta\to 0}m(x,\delta)$，其中 $m(x,\delta)$ 为函数 $f(x)$ 在点 x 的 δ 邻域上取值的下确界	
$\chi_A(x)$	集合的特征函数	characteristic function of a set	$\chi_A(x)=\begin{cases}1 & (x\in A),\\ 0 & (x\notin A)\end{cases}$	
ap $\overline{\lim}$	近似上极限	approximate upper limit	ap $\overline{\lim\limits_{x\to x_0}}f(x)=\inf\limits_{E}\overline{\lim\limits_{x\to x_0}}{}_E f(x)$	
ap $\underline{\lim}$	近似下极限	approximate lower limit	ap $\underline{\lim\limits_{x\to x_0}}f(x)=(\sup\limits_{E}\underline{\lim\limits_{x\to x_0}}{}_E f(x))$	
ap lim	近似极限	approximate limit	ap $\lim\limits_{x\to x_0}f(x)$ 表示 ap $\overline{\lim\limits_{x\to x_0}}f(x)=$ ap $\underline{\lim\limits_{x\to x_0}}f(x)$	
$(L)\int_E f(x)\mathrm{d}x$	勒贝格积分	Lebesgue integral	若 $f(x)$ 是可测集 $E\subset\mathrm{R}^n$ 上的 (L) 可测函数，则称 $(L)\int_E f(x)\mathrm{d}x$ 为勒贝格积分	简称 L 积分
$D^-f(x_0)$	左上导数	left upper derivative	$D^-f(x_0)=\overline{\lim\limits_{\xi\to x_0^-}}\frac{f(\xi)-f(x_0)}{\xi-x_0}$	
$D_-f(x_0)$	左下导数	left lower derivative	$D_-f(x_0)=\underline{\lim\limits_{\xi\to x_0^-}}\frac{f(\xi)-f(x_0)}{\xi-x_0}$	
$D^+f(x_0)$	右上导数	right upper derivative	$D^+f(x_0)=\overline{\lim\limits_{\xi\to x_0^+}}\frac{f(\xi)-f(x_0)}{\xi-x_0}$	
$D_+f(x_0)$	右下导数	right lower derivative	$D_+f(x_0)=\underline{\lim\limits_{\xi\to x_0^+}}\frac{f(\xi)-f(x_0)}{\xi-x_0}$	
$\ll$	绝对连续	absolute continuity	$\gamma\ll\mu$ 表示广义测度 γ 关于 μ 是绝对连续的. 即当 $\lvert\mu\rvert(A)=0$ 时有 $\gamma(A)=0$，其中 $\lvert\mu\rvert$ 是 μ 的全变差	
$\perp$	相互奇异	mutually singular	$\gamma\perp\mu$ 表示 γ 与 μ 是相互奇异的，即存在两个不相交的可测集 A 与 B 使得 $\Omega=A\cup B$，且对任意可测集 E，有 $\lvert\mu\rvert(A\cup E)=\lvert\gamma\rvert(B\cap E)=0$，其中 $\lvert\gamma\rvert,\lvert\mu\rvert$ 分别是 γ 和 μ 的全变差	
$(\Gamma)\int_\Omega x(t)\mathrm{d}\mu$	盖尔范德积分	Gelfand integral	设 $x(t)$ 为 Ω 到巴拿赫空间 X 的向量函数，若对 $\forall f\in X^*$，当 $f(x(t))$ 在 Ω 上可积时必存在 $x^{**}\in X$ 使 $x^{**}=\int_\Omega f(x(t))\mathrm{d}\mu$，则称 x^{**} 为盖尔范德积分	亦称盖尔范德意义下的弱 * 积分

符号	中文名称	英文名称	意义或举例	备注
$(P)\int_A x(t)\mathrm{d}\mu$	佩蒂斯积分	Pettis integral	若$\int_A f(x(t))\mathrm{d}\mu = f(x_A)$，则$(P)\int_A x(t)\mathrm{d}\mu = x_A$	亦称弱积分
$(B)\int_\Omega x(t)\mathrm{d}\mu$	博赫纳积分	Borchner integral	1.若 $x(t)$ 是 Ω 上可测函数，则 $(B)\int_\Omega x(t)\mathrm{d}\mu = \sum_{k=1}^{\infty} x_k\mu(A_K)$； 2. 对于一般的强可测函数 $x(t)$，则 $(B)\int_\Omega x(t)\mathrm{d}\mu = \lim_{n\to\infty}(B)\int_\Omega x_n(t)\mathrm{d}\mu$.	
$(BK)\int_\Omega x(t)\mathrm{d}\mu$	伯克霍夫积分	Birkhoff integral	$(BK)\int_\Omega x(t)\mathrm{d}\mu = \bigcap_\Delta J(x,\Delta)$，其中 $J(x,\Delta)$ 是 $\{\sum_{i=1}^{\infty}\mu(A_i)x(t_i) \mid t_i\in A_i\}$ 的凸闭包	
$\mathscr{U}_g^*(E)$	(L-S)外测度	(L−S)outer measure	$\mathscr{U}_g^*(E)=\inf\{\sum_{K\geqslant 1}\mathscr{U}_g(I_k) \mid \{I_k\}$ 为可数个覆盖 E 的左开右闭区间$\}$	
$\mathscr{U}_g(E)$	(L-S)测度	(L−S)measure	当任意点集 T 能分解成 E 内部分 $T\cap E'$ 和 E 外部分 $T\cap E^c$ 时，相应的(L-S)外测度具有可加性，则 E 称为 $g(x)$的(L-S)可测集，此时外测度 $\mathscr{U}_g^*(E)$就称为 E 的由分布函数$g(x)$引出的(L-S)测度	
$(L\text{-}S)\int_E$	(L-S)积分	(L-S)integral	$\int_E f(x)\mathrm{d}g(x)=\int_E f^+(x)\mathrm{d}g(x)-\int_E f^-(x)\mathrm{d}g(x)$，其中 $f^+(x)$，$f^-(x)$分别为 $f(x)$正部和负部，且至少有一个有极限	(L-S)积分是勒贝格-斯蒂尔切斯积分的简称
$D(*)\int_a^b$	狭义当茹瓦积分	Denjoy integral in the restricted sense	$(D(*))\int_a^b f(x)\mathrm{d}x = F(b)-F(a)$，其中 $F(x)$ 是狭义一般绝对连续函数，且在$[a,b]$上 $F'(x)=f(x)$ a.e.	狭义当茹瓦积分是勒贝格积分和黎曼积分的一种推广
$D_{ap}f(x_0)$	近似导数	approximate derivative	$D_{ap}f(x_0) = \mathrm{ap}\lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}$	
$\underline{D}_{ap}f(x_0)$	近似下导数	approximate lower derivative	$\underline{D}_{ap}f(x_0) = \mathrm{ap}\varliminf_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}$	
$\overline{D}_{ap}f(x_0)$	近似上导数	approximate upper derivative	$\overline{D}_{ap}f(x_0) = \mathrm{ap}\varlimsup_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}$	
Π_K	庞特里亚金空间	Pontrjagin space	设 $H=H_-\oplus H_+$ 是正则分解，$\dim H_\pm = k<+\infty$，称$(H,[\cdot,\cdot])$为具有正(负)指标的庞特里亚金空间	
π	克莱因空间	Klein space	设 $H=H_-\oplus H_+$ 是正则分解，$\dim H_\pm = +\infty$，称$(H,[\cdot,\cdot])$为克莱因空间	
$\rho(T)$	正则集	Regular set	设 T 是空间 X 的线性算子，如果 $\lambda I-T$ 是正则算子，那么称 λ 为 T 的正则点. 复平面上正则点全体称为正则集	亦称豫解集
$\sigma(T)$或$\mathrm{sp}(T)$	谱集	spectrum	$\rho(T)$ 的余集 $C\backslash\rho(T)$. $\sigma_P(T)$，$\sigma_a(T)$，$\sigma_r(T)$，$\sigma_c(T)$ 分别表示点谱、近似点谱、剩余谱、连续谱	
$\deg(T,\Omega,P)$	拓扑度	topological degree	映射T在区域Ω上关于P点的拓扑度是一个整数，它是方程 $T(x)=P$ 在 Ω 中解的“代数个数”的某种稳定的度量	
$F((x))$	形式幂级数域	domain of formal power series	由 F 上关于X的形式幂级数 $\alpha(x)=q_rx^r+q_{r+1}x^{r+1}+\cdots$ $(q_r\neq 0, r\in Z)$ 按照通常加、乘运算组成一个域	
$\delta(x)$	狄拉克δ函数	Dirac δ-function	$\delta(x)=\begin{cases}+\infty & (x=0),\\ 0 & (x\neq 0).\end{cases}$	
$e\subset(A)$	平衡包	equilibrium hull	包含 A 的最小平衡集称为 A 的平衡包	
$(P)\int_a^b f(x)\mathrm{d}x$	佩龙积分	Perron integral	$(P)\int_a^b f(x)\mathrm{d}x=\inf\{U(b)\}=\sup\{V(b)\}$，其中$U(x)$和 $V(x)$ 分别是 $f(x)$ 在$[a,b]$上的上函数和下函数	$f(x)$在$[a,b]$上的佩龙积分值和勒贝格积分值相等
$(W)\int_a^b f(x)\mathrm{d}x$	瓦尔德积分	Wald integral	$(W)\int_a^b f(x)\mathrm{d}x=\sup_G(G(b))-(G(a))=\inf(H(b)-H(a))$，其中 $H(x)$，$G(x)$ 各为 $f(x)$ 的瓦尔德上、下函数	瓦尔德积分与佩龙积分等价

符 号	中文名称	英文名称	意 义 或 举 例	备 注
$(H)\int_a^b f(x)\mathrm{d}x$	亨斯托克积分	Henstock integral	一种定积分，亨斯托克积分包括(R)积分，也包括(L)积分	
$(M)\int_a^b f(x)\mathrm{d}x$	马克仙积分	Mcshane integral	一种定积分，马克仙积分与勒贝格积分等价	
$f_n \xrightarrow{L^p} f$	L^p的强收敛	strong convergence in L^p	若$f_n(x), f(x) \in L^p(E), (1 \leqslant p < +\infty, n = 1, 2, \cdots,)$，且存在$\Vert f_n - f \Vert_p \to 0(n \to \infty)$，则称$\{f_n(x)\}$强收敛于$f(x)$	亦称按L^p范数收敛于$f(x)$
$f_n \xrightarrow{W} f$	L^p的弱收敛	weak convergence in L^p	若$f_n(x), f(x) \in L^p(E), g(x) \in L^q(E), (1 < p, q < +\infty, n = 1, 2, \cdots,)$且$1/p + 1/q = 1$时，$\lim\limits_{n\to\infty}\int_E f_n(x)g(x)\mathrm{d}x = \int_E f(x)\mathrm{d}x$成立，则称$\{f_n(x)\}$弱收敛于$f(x)$	
l^p	l^p空间	l^p space	所有满足$\Vert x \Vert_p = (\sum\limits_{k=1}^{\infty} \vert x_k \vert^p)^{1/p} < +\infty$的数列$x$组成之集	
l^∞	l^∞空间	l^∞ space	满足$\vert x_n \vert \leqslant M < +\infty (n = 1, 2, \cdots,)$的所有数列之集．$x$的范数由$\Vert x \Vert_\infty = \sup\limits_n \{\vert x_n \vert\}$定义	
$\Lambda(\psi)$	洛伦茨空间	Lorentz space	$\Lambda(\psi) = \{f \in S[0,1] \mid \Vert f \Vert < +\infty\}$称为洛伦茨空间	
L_Φ^*	奥尔里奇空间	Orlicz space	所有使得$\Vert f \Vert = \inf\left\{\lambda > 0 \middle\vert \int_{\mathbb{R}^n} \Phi(\lambda^{-1}\vert f(t) \vert)\mathrm{d}t \leqslant 1\right\} < +\infty$成立的$\mathbb{R}$上的可测函数$f$之集	
ent	拓扑熵	toplogical entropy	这是用于拓扑动力学中的一个概念	
$J_n^{(\alpha,\beta)}(x)$	雅可比多项式	Jacobi polynomials	$[-1,1]$上关于权$\omega(x) = (1-x)^\alpha(1+x)^\beta$的正交多项式 $J_n^{(\alpha,\beta)}(x) = \frac{1}{n!2^n\omega(x)\mathrm{d}x^n}\frac{\mathrm{d}^n}{\mathrm{d}x^n}[(x^2-1)^n\omega(x)]$ $(n = 1, 2, \cdots)$	
$r_n(x)$	拉德马赫尔函数	Rademacher functions	$r_n(x) = \mathrm{sig}\ n\sin 2^{n+1}x \quad (0 \leqslant x \leqslant 1, n = 1, 2, \cdots)$	
$W_n(x)$	沃尔什函数	Walsh functions	$W_n = r_{k_1}(x)r_{k_2}(x)\cdots r_{k_p}(x) \quad (0 \leqslant x \leqslant 1)$	
$B_n(f,x)$	伯恩施坦多项式	Bernstein polynomial	$B_n(f,x) = \sum\limits_{k=0}^{n}\binom{n}{k}x^x(1-x)^{n-k}f\left(\frac{k}{n}\right)$	亦称伯恩施坦算子
$H_\varepsilon(A)$	度量熵	metric entropy	设A是巴拿赫空间X的紧子集，A的ε覆盖$\{U_k\}_{k=1}^n$，令$N_\varepsilon(A) = \min n$，则$H_\varepsilon(A) = \log N_\varepsilon(A)$	
$H_\varepsilon^X(A)$	A关于X的熵	entropy of A with respect to X	设A是巴拿赫空间X的紧子集，A的ε网$\{x_k\}_{k=1}^p$，令$P_\varepsilon(A) = \min P$，则$H_\varepsilon^X(A) = \log P_\varepsilon(A)$	
$C_\varepsilon(A)$	容量	capacity	设A是巴拿赫空间X的紧子集，A的$\varepsilon\gamma$分离$\{y_k\}_{k=1}^m$，令$M_\varepsilon(A) = \max m$，则$H_\varepsilon(A) = \log M_\varepsilon(A)$	
L_n	勒贝格常数	Lebesgue constant	$L_n = \frac{4}{\pi^2}\log(n+1) + o(1)$	
$\deg(\pi)$	分歧阶	ramification order	使π在A_k恒为1的最小整数k	
PX	X的子集簇	subsets of X	集合X的一切子集组成的集合	亦称幂集合
Δ	对称差	symmetric difference	$A\Delta B$的对称差指属于A但不属于B，或属于B但不属于A的一切元素组成的集合	
$P \cdot P \cdot P$	近乎处处	approximately everywhere	设$P = P(x)$是一个与x无关的性质，如果使P不成立的点全体所成之集A为零内容集，则称P是近乎处处成立的	
$q \cdot P \cdot$	拟乎处处	quasi-everywhere	设$P = P(x)$是一个与x无关的性质，如果A为零外容集，则称P是拟乎处处成立的	
$\mathrm{cap}(G)$	χ容量	χ-capacity	对于相对紧的开集G，记$\mathrm{cap}(G) = \int \mathrm{d}\sigma_G$，其中$\sigma_G$是由$R_{\omega x}^G = \chi^* \sigma_G$所确定的惟一测度	
U_K^μ	位势	potential	测度μ的K位势为 $U_K^\mu = \int_\Theta K(x,y)d\mu(y) \quad (x \in \Omega)$	
U_α^μ	里斯位势	Riesz potential	对于位势U_K^μ，当$\Omega = R^n(n \geqslant 3), 0 < \alpha < n, \kappa(x,y) = \vert x - y \vert^{\alpha-n}$时，称为里斯位势	

符号	中文名称	英文名称	意义或举例	备注
U_ξ^g	牛顿位势	Newtonian potential	对于里斯位势 $\alpha=2$ 时，称为牛顿位势	
(f,g)	内积	inter product	$(f,g)=\int_\Omega f(x)g(x)\mathrm{d}u(x)$	
σ	舒伯特符号	Schubert symbol	$\sigma=(\sigma_1,\sigma_2,\cdots,\sigma_n)$ 表示 n 个整数组成的一个序列，其中 $1\leqslant\sigma_1<\sigma_2<\cdots<\sigma_n\leqslant m$	
Lin E	线性包	linear hull	$\mathrm{Lin}\,E=\{x\mid x=\sum_{y\in E}\lambda_y y,\lambda_y\in R$，有限个不为零$\}$	$\mathrm{Lin}E$ 亦表示凸集 E 的支撑子空间
affE	仿射包	affine hull	$\mathrm{aff}E=\{x\mid x=\sum_{y\in E}\lambda_y y,\lambda_y\in R$，有限个不为零，$\sum_{y\in E}\lambda_y=1\}$	
coneE	锥包	cone hull	$\mathrm{cone}E=\{x\mid x=\lambda y,y\in E,\lambda>0\}=\bigcup_{\lambda>0}\lambda E$	
coE	凸包(凸集)	covex hull	$\mathrm{co}E=\{x\mid x=\sum_{y\in E}\lambda_y y,\lambda_y\in[0,1]$，有限个不为零，$\sum_{y\in E}\lambda_y=1\}$	
clcoE	闭凸包	closed convex hull	以 $\dot{C}$ 为内集的全体闭包凸集的交	
epif	上图	epigraph	$\mathrm{epi}f=\{(x,a)\in X\times R\mid f(x)\leqslant a\}$	
K	核	kernel	$C\subset R^n,\forall\ y\in C,0\leqslant\lambda\leqslant1$，满足$(1-\lambda)x+\lambda y\in C$ 的全体 $x\in C$ 的集合称为 C 的核	
exp C	暴露点集	exposing point set	C 的全体暴露点的集合	
ext C	极点集	extreme point set	C 的全体极点的集合	
$f'(x:y)$	单边方向导数	one-side directional derivative	$f'(x:y)=\lim\limits_{\lambda\downarrow0}\dfrac{f(x+\lambda y)-f(x)}{\lambda}$	
$\partial f(x)$	次微分	subdifferential	$f(x)$ 在 X 的次梯度的全体	
$I_V(M)$	奇点的指标	index of critical points	V 的孤立奇点 M 沿曲线 C_r 的旋转数	
u. a. p.	一致概周期函数	uniformly almost periodic functions	设 $f(t,x)\in C(\mathrm{R}\times D,E^n)$，$S$ 是 D 的紧集，若对任给序列$\{a'_n\}$，存在子序列$\{a_n\}\subset\{a'_n\}$，使 $T_af(t,x)=\lim\limits_{n\to\infty}f(t+a_n,x)$ 在 $R\times S$ 上一致地成立，则称 $f(t,x)$ 是一致概周期函数，$x\in D$	
a. a. p.	渐进概周期函数	asymptotically almost periodic functions	如果 $\varphi(t)$ 有分解式 $\varphi(t)=p(t)+q(t)$，其中 $p(t)$ 是 R 上的概周期函数，$q(t)$ 是定义在 R^+（或 R^-）上的连续函数，当 $t\to+\infty$ 或 $(t\to-\infty)$ 时有 $q(t)\to0$，则称 $\varphi(t)$ 是 R^+（或 R^-）上的渐进概周期函数	
RFDE(f)	滞后型泛函微分方程	retarded function differential equation	$\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=f(t,x(t),x(t-h_1),\cdots,x(t-h_m))$. $(h_1,h_2,\cdots,h_m$ 是正定数，$h_1<h_2<\cdots<h_m)$	RFDE 是英文名中四个单词的第一个字母
H. S.	哈密顿系统	Hamilton's system	指形如$\dfrac{\mathrm{d}p}{\mathrm{d}t}=-\dfrac{\partial H}{\partial q},\dfrac{\mathrm{d}q}{\mathrm{d}t}=\dfrac{\partial H}{\partial p},H=H(p,q,t)$ 的一阶偏微分方程	亦称典型系统或正则系统
$\int^x a(s)\mathrm{d}s$	反导数	antiderivative	表示 $a(x)$ 的反导数	

概率统计(Probability & Statistics)

符号	中文名称	英文名称	意义或举例	备注
P、P_r	概率	probability	$P(E)$表示事件 E 的概率，$P_r(\xi)$表示事件 ξ 的概率	$P_{n,m}$ 表示在 n 次独立实验中出现 m 次事件的概率
$P(\mid)$	条件概率	conditional probability	$P(A\mid B)$表示发生了事件 B 的条件下，事件 A 的概率	
E,M	期望（或均值）	expectation (or mean)	$E\xi$,$M\xi$ 表示随机变量 ξ 的期望(或均值)	亦可记为 $E(\xi)$，$M(\xi)$
D、σ^2	方差	variance	$D\xi$，$\sigma^2\xi$ 表示随机变量 ξ 的方差	亦可记为 $D(\xi)$，$\sigma^2(\xi)$，$\mathrm{Var}\xi$

符号	中文名称	英文名称	意义或举例	备注
cov	协方差	covariance	$\mathrm{cov}(\xi,\eta)$表示随机变量ξ和η的协方差	或记为$\sigma_{\xi,\eta}$
$E(\vert)$, $M(\vert)$	条件期望（或条件均值）	conditional expectation or conditional mean	$E(\xi\vert y)$，$M(\xi\vert y)$表示随机变量ξ关于条件y的条件期望（或均值）	
ρ, r	相关系数	correlation coefficient	$\rho(\xi,\eta)$，$\rho_{\xi,\eta}$，$r(\xi,\eta)$表示随机变量ξ和η的相关系数	在不致误会时，亦可记为ρ或r
Ω	基本事件空间	elementary event space	Ω是由n个基本事件$\omega_i(i\in\mathrm{N})$构成的基本事件空间，即$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$	
$F_n(\ )$	频率	frequency	频率$F_n(A)$等于频数$f_n(A)$与试验总次数n之比，即 $F_n(A)=\dfrac{f_n(A)}{n}$	
$F(\vert)$	条件分布函数	conditional distribution function	ξ和η为随机变量，则称$F(y\vert x)$为在$\xi=x$条件下η的条件分布函数	
ν_k	k阶原点矩	origin moment of the k-th order	ξ的k阶原点矩$\nu_k=E(\xi^k)$	
μ_k	k阶中心矩	central moment of the k-th order	ξ的k阶中心矩$\mu_k=E(\xi-E\xi)^k$	
α_k	k阶原点绝对矩	origin absolute moment of the$k$-th order	ξ的k阶原点绝对矩$\alpha_k=E\vert\xi\vert^k$	
β_k	k阶中心绝对矩	central absolute moment of the k-th order	ξ的k阶中心绝对矩$\beta_k=E\vert\xi-E\xi\vert^k$	
$E(\ )$	混合矩	mixed moment	若$E\vert\xi^k\eta^l\vert<\infty$，$k,l\in\mathrm{N}$，则称$E(\xi^k\eta^l)$为$\xi$和$\eta$的$k+l$阶混合矩	
$E[\ ]$	中心混合矩	central mixed moment	若$E(\vert\xi-E\xi\vert^k\vert\eta-E\eta\vert^l)<\infty$，且$k,l\in\mathrm{N}$，则称$E[(\xi-E\xi)^k(\eta-E\eta)^l]$为$\xi$和$\eta$的$k+l$阶中心混合矩	
$B(n,p)$	二项分布	binomial distribution	分布列为$b(k;n,p)=\binom{n}{k}p^kq^{n-k}$ $(0<p<1,q=1-p,k=0,1,2,\cdots,n)$	
$NB(m,p)$	负二项分布	negative binomial distribution	密度函数为$p_x=\Gamma(m+x)[\Gamma(m)x!]^{-1}p^mq^x$（$m$为整数，$0<p<1$，$q=1-p,x=0,1,2,\cdots$）	
$G(p)$ 或 $g(k;p)$	几何分布	geometric distribution	密度函数为 $p_x=pq^x(0<p<1,q=1-p,x=0,1,2,\cdots)$	
$H(N,n,p)$	超几何分布	hypergeometric distribution	密度函数为$p_x=\binom{Np}{x}\binom{Nq}{n-x}\Big/\binom{N}{n}$（$x$为整数，$N,Np$，$n$为正整数，$N\geqslant n$，$0\leqslant x\leqslant Np$，$0\leqslant n-x\leqslant Nq$，$0<p<1$，$q=1-p$）	
$M(n;p_1,\cdots,p_{k+1})$	多维超几何分布	multiple hypergeometric distribution	密度函数为$p_{x_i}=\binom{Np_1}{x_1}\cdots\binom{Np_{k+1}}{x_{k+1}}\Big/\binom{N}{n}$ （$i=1,2,\cdots,k+1$），$x_1,x_2,\cdots,x_{k+1}$是整数，$N,Np_1,\cdots,Np_{k+1},n$是正整数，$x_{k+1}=n-(x_1+\cdots+x_k)$，$p_1+p_2+\cdots+p_{k+1}>0$）	
$P(\lambda)$ 或 $P(k;\lambda)$	泊松分布	Poisson distribution	分布列为$p(k;\lambda)=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}(\lambda>0,k=0,1,2,\cdots)$	
$U(a,b)$ 或 $U[a,b]$	均匀分布	uniform distribution	密度函数为$p(x)=\begin{cases}1/(b-a) & (a\leqslant x\leqslant b),\\0 & (其他),\end{cases}$ 其中$a<b$为常数	
$N(\mu,\sigma^2)$	正态分布	normal distribution	密度函数为$p(x)=\dfrac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$， $(-\infty<x<+\infty,\sigma>0,\mu$为常数)	亦称高斯分布
$C(\lambda,\mu)$	柯西分布	Cauchy distribution	密度函数为$p(x)=\dfrac{1}{\pi}\cdot\dfrac{\lambda}{\lambda^2+(x-\mu)^2}$，其中$x$为实数，$\lambda>0$，$\mu$为常数	
$\Gamma(\lambda,r)$	伽马分布	gamma distribution	密度函数为$p(x)=\begin{cases}\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}\mathrm{e}^{-\lambda x} & (x>0),\\0 & (x\leqslant 0),\end{cases}$ 其中$r>0,\lambda>0$为常数	亦可记为$G(\lambda,r)$

数学符号表

符号	中文名称	英文名称	意义或举例	备注
$e(\lambda)$	指数分布	exponential distribution	密度函数为 $p(x)=\begin{cases}\lambda e^{-\lambda x} & (x\geqslant 0),\\ 0 & (x<0),\end{cases}$ 其中 λ 为常数	亦可记为 $e(\mu,\sigma)$
$W(\lambda,\alpha)$	韦布尔分布	Weibull's distribution	密度函数为 $p(x)=\begin{cases}\alpha\lambda x^{\alpha-1}\exp(-\lambda x^{\alpha}) & (x>0),\\ 0 & (x\leqslant 0),\end{cases}$ 其中 $\lambda>0$，$\alpha>0$ 为常数	
$\chi^2(n)$	χ^2 分布	Chi-square distribution	密度函数为 $p(x)=\begin{cases}\dfrac{x^{(n-2)/2}e^{-x/2}}{2^{n/2}\Gamma\left(\dfrac{n}{2}\right)} & (x>0),\\ 0 & (x\leqslant 0),\end{cases}$ 其中 n 为正整数	
$\mathrm{Ln}(\mu,\sigma^2)$	对数正态分布	logarithmic normal distribution	密度函数为 $P(x)=\begin{cases}\dfrac{1}{\sigma x\sqrt{2\pi}}e^{-(\ln x-a)^2/2\sigma^2} & (x>0),\\ 0 & (x\leqslant 0),\end{cases}$ 其中 $a,\sigma>0$ 为常数	
$t(n)$	学生分布	Student's distribution	密度函数为 $p(x)=\dfrac{1}{\sqrt{n\pi}}\dfrac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)}\left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$， 其中 n 为正整数	亦称 t 分布
$F(n_1,n_2)$	F 分布	F-distribution	密度函数为 $p(x)=\begin{cases}\dfrac{x^{\frac{n_1}{2}-1}}{B\left(\dfrac{n_1}{2},\dfrac{n_2}{2}\right)}n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}(n_2+n_1x)^{-\frac{n_1+n_2}{2}} & (x>0),\\ 0 & (x\leqslant 0),\end{cases}$ 其中 n_1,n_2 为正整数	
$E(\alpha,\beta)$	极值分布	extremal distribution	密度函数为 $p(x)=\dfrac{1}{\beta}\exp\left\{\exp\left(-\dfrac{x-\alpha}{\beta}\right)-\dfrac{x-\alpha}{\beta}\right\}$， 其中 x,α 均为实数，β 为常数	
$\chi^2(n,\lambda)$	非中心 χ^2 分布	non-central chi-square distribution	密度函数为 $p(x)=\begin{cases}\dfrac{\exp\left\{-\left(\dfrac{x+\lambda}{2}\right)\right\}}{2^{n/2}}\sum\limits_{j=0}^{\infty}\dfrac{x^{\frac{n}{2}+j-1}\lambda^j}{\Gamma\left(\dfrac{n}{2}+j\right)2^{2j}j!} & (x>0),\\ 0 & (x\leqslant 0),\end{cases}$ 其中 n 为自由度；$\lambda>0$ 为非中心参数	
$t(n,\delta)$	非中心 t 分布	non-central t-distribution	密度函数为 $p(x)=\dfrac{n^{n/2}\exp(-\delta^2/2)}{\sqrt{\pi}\,\Gamma\left(\dfrac{n}{2}\right)(n+x^2)^{(n+1)/2}}\sum\limits_{m=0}^{\infty}\Gamma\left(\dfrac{n+m-1}{2}\right)\left(\dfrac{\delta^m}{m!}\right)\left(\dfrac{2x^2}{2+x^2}\right)^{\frac{n}{2}}$， 其中 n 为自由度，δ 为实数，且是非中心参数	
$F(m,n;\lambda)$	非中心 F 分布	non-central F-distribution	密度函数为 $p(x)=\begin{cases}\dfrac{m^{\frac{m}{2}}n^{\frac{n}{2}}}{\Gamma\left(\dfrac{n}{2}\right)}e^{-\frac{\lambda}{2}}x^{\frac{m}{2}-1}\sum\limits_{k=0}^{\infty}\dfrac{\left(\dfrac{\lambda m x}{2}\right)^k\Gamma\left(\dfrac{m+n}{2}+k\right)}{\Gamma\left(\dfrac{m}{2}+k\right)k!(mx+n)^{\frac{m+n}{2}+k}} & (x>0),\\ 0 & (x\leqslant 0),\end{cases}$ 其中 m,n 为二自由度，λ 为非中心参数	
$X_1^{(n)}$	最小顺序统计量	smallest order statistics	$X_1^{(n)}=\min\limits_{1\leqslant i\leqslant n}X_i$ 表示样本观察值中最小者	
$X_n^{(n)}$	最大顺序统计量	largest order statistics	$X_n^{(n)}=\max\limits_{1\leqslant i\leqslant n}X_i$ 表示样本观察值中最大者	
$\bar{x}$	样本均值	sample mean	$\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i=\int_{-\infty}^{\infty}x\mathrm{d}F_n(x)$	
s^2	样本方差	sample variance	$s^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2=\int_{-\infty}^{\infty}(x-\bar{x})^2\mathrm{d}F_n(x)$	
a_k	样本 k 阶原点矩	sample origin moment of the k-th order	$a_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i^k=\int_{-\infty}^{\infty}x^k\mathrm{d}F_n(x)\quad(k=2,3,\cdots)$	

符号	中文名称	英文名称	意义或举例	备注
b_k	样本 k 阶中心矩	sample central moment of the k-th order	$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k=\int_{-\infty}^{\infty}(x-\bar{x})^k\mathrm{d}F_n(x)$ $(k=2,3,\cdots)$	
μ	总体均值	population mean	$\mu=E(X)$	
σ^2	总体方差	population variance	$\sigma^2=D(X)=E(X-\mu)^2$	
α_k	总体 k 阶原点矩	population origin moment of the k-th order	$\alpha_k=E(X^k)=\int_{-\infty}^{\infty}x^k\mathrm{d}F(x)$	
μ_k	总体 k 阶中心矩	population central moment of the k-th order	$\mu_k=E(X-\mu)^k=\int_{-\infty}^{\infty}(x-\mu)^k\mathrm{d}F(x)$	
Md	样本中位数	sample median	$\mathrm{Md}X=\begin{cases}X_{k+1}, & 若\ n=2k+1,\\ (X_k+X_{k+1})/2, & 若\ n=2k\end{cases}$	亦可用 $\tilde{X}$ 表示
Sk	样本偏度	sample skewness	样本三阶中心矩除以样本二阶中心矩的3/2次幂的商，即 $\mathrm{Sk}=\frac{b_3}{(b_2)^{3/2}}$	亦称样本偏态或偏态系数
Kur	样本峰度	sample kurtosis	样本四阶中心矩除以样本二阶中心矩的平方再减去3，即 $\mathrm{Kur}=\frac{b_4}{(b_2)^2}-3$	亦称样本峭度
$\mathrm{d}f$，f	自由度	degree of freedom	$\mathrm{d}f_A$，f_A 表示因素 A 的自由度	
$E_x(s)$	特征函数	characteristic	函数 $\mathrm{e}^{\mathrm{i}sX}$ 的数学期望，即 $E_x(s)=M[\mathrm{e}^{\mathrm{i}sX}]$	
$H[x]$	熵	entorpy	离散型随机变量 x 的熵 $H[x]=-\sum_{i=1}^{n}P_i\log_aP_i$；连续型随机变量 x 的熵 $H[x]=-\int_{-\infty}^{+\infty}f(x)\log_af(x)\mathrm{d}x$	
$f(k;r,p)$	帕斯卡分布式	Pascal distribution	分布函数为 $f(k;r,p)=C_{k-1}^{r-1}p^rq^{r-1}\quad(k=r,r+1,\cdots)$	
$P_{i.}$ 或 $P_{.j}$	边缘概率	boundary probability	离散型随机变量的边缘概率分布式为 $P_{i.}=\sum_jP_{ij}$，$P_{.j}=\sum_iP_{ij}$	
$N(\mu,\Sigma)$ 或 $N_n(\mu,\Sigma)$	多维正态分布	normal distribution	N 维正态分布的密度函数为 $\varphi(x)=\frac{1}{(2\pi)^{n/2}\sqrt{\lvert\Sigma\rvert}}\exp\{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\}(x\in\mathbf{R}^n)$	
S_n^*	S_n 的标准化	standardization of S	$S_n^*=\sum_{k=1}^{n}(X_k-a_k)/S_n$	
ω	样本点	sample point	随机试验的每一个可能的结果	亦称基本事件
ϕ	不可能事件	non-probability event	随机试验不可能发生的结果	
E^n	伯努利试验	Bernoulli trials	随机试验 E 只有两个可能的结果，并且其概率为 p，q，其中 $q=1-p$，把 E 独立地重复 n 次试验构成了一个试验	亦称伯努利概型
$\sigma\xi$	标准差	root-mean square deviation	方差的平方根	亦称根方差
CL	中线	middle line	表示控制图中中线	
UCL	上控制线	upper control linear	表示控制图中上控制线	
LCL	下控制线	lower control linear	表示控制图中下控制线	
$(n\mid C)$	抽检方案	sampling inspection plan	表示子样的容量为 n 和允许的不合格数为 C	
T	寿命	longevity	对任一特定个体(产品或生命体)，从某个标准时间起在规定时间 t 内失效(或死亡)	
$R(t)$	可靠度	reliability	产品在规定的条件下，规定的时间内，完成规定功能的概率	
ρ_r	可靠寿命	reliability life	使可靠度等于给定值 r 的时间	$\rho_{0.5}$称为中位寿命
$\lambda(t)$	失效率	failure rate	产品工作到 t 时刻后单位时间内发生失效的概率	
MTBF	平均无故障工作时间	mean time between failures	平均寿命对可修复产品	

符号	中文名称	英文名称	意义或举例	备注
MTTF	失效前的平均工作时间	worked mean time before failure	平均寿命对不可修复产品	
PDF	概率分布函数	probability distribution function	$F(x)=P(\xi(\omega)<x), x\in(-\infty,+\infty)$	简称分布函数
MLE	极大似然估计	maximum likelihood estimate	使似然函数 $L(p)$ 达到极大值的参数 P	
$\hat{\theta}$	估计量	estimator	当区间 $(\hat{\theta}_1,\hat{\theta}_2)$ 以某一指定的概率包含 θ 时，称 $(\hat{\theta}_1,\hat{\theta}_2)$ 为函数 θ 的区间估计	
R	样本极差	sample range	$R=\max\{x_1,x_2,\cdots,x_n\}-\min\{x_1,x_2,\cdots,x_n\}$ 表示取样本中最大值与最小值之差	亦称样本范围，又称样本全距
H_0	原假设	null hypothesis	假设检验中，对有关总体需要作出判断的待检验的命题的假设	亦称零假设
H_1, H_a	备择假设	alternative hypothesis	假设检验中，异于原假设的另一假设	亦称择一假设
u,λ,t	临界值	critical value	u_a，λ_a，t_a 表示置信度为 α 的临界值	
Q	离差平方和	sum of squares of deviations	总离差平方和 $Q=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x})^2$； 组内离差平方和 $Q_1=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2$； 组间离差平方和 $Q_2=n\sum_{i=1}^{m}(\bar{x}_i-\bar{x})^2$； 因素 A 的离差平方和 $Q_A=n\sum_{i=1}^{n}(\bar{x}_{i\cdot}-\bar{x})^2$； 误差平方和 $Q_E=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x})^2$	
*	显著性标记	significance marked	* 表示作用显著，* * 表示作用高度显著	
×	交互作用	interaction	$A\times B$ 表示因素 A,B 的交互作用	
$L(\ )$	正交表示标记	orthogonal layout marked	$L_4(2^3)$ 表示二水平三因素，需作四次试验的正交表示	
vec	列拉直算子	operator of according to columns draw line	将矩阵 $A=(a_{ij})_{n\times m}$ 中的元按列依次拉直排序，即 $\mathrm{vec}(A)=(a_{11},a_{21},\cdots,a_{n1},a_{12},a_{22},\cdots,a_{n2},\cdots,a_{nm})$	
ran	行拉直算子	operator of according to rowsdraw line	将矩阵 $A=(a_{ij})_{n\times m}$ 中的元按行依次拉直排序，即 $\mathrm{ran}(A)=(a_{11},a_{12},\cdots,a_{1m},a_{21},a_{22},\cdots,a_{2m},\cdots,a_{nm})$	

应用数学（Applied mathematics）

符号	中文名称	英文名称	意义或举例	备注
$\underset{\sim}{A}$	模糊子集	fuzzy subset	$\underset{\sim}{A}=\{x,\mu_{\underset{\sim}{A}}(x)\mid x\in X\}$，其中集 X 为论域，$\forall\ x\in X$，$\mu_{\underset{\sim}{A}}(x)\in[0,1]$ 是模糊子集 $\underset{\sim}{A}$ 的隶属函数	亦称模糊集、弗晰集、不分明集、乏晰集等
$\vee$	模糊子集的上确界	supremum of fuzzy subset	若 $\{a_t\mid t\in T\}$ 是实数集，则 $\bigvee_{t\in T} a_t=\sup\{a_t\mid t\in T\}$	
$\wedge$	模糊子集的下确界	infimum of fuzzy subset	若 $\{a_t\mid t\in T\}$ 是实数集，则 $\bigwedge_{t\in T} a_t=\inf\{a_t\mid t\in T\}$	
$\overset{\wedge}{+}$	代数和	algebraic sum	$\mu_{\underset{\sim}{A}\cup\underset{\sim}{B}}(x)=\mu_{\underset{\sim}{A}}(x)\overset{\wedge}{+}\mu_{\underset{\sim}{B}}(x)$ $=\mu_{\underset{\sim}{A}}(x)+\mu_{\underset{\sim}{B}}(x)-\mu_{\underset{\sim}{A}}(x)\mu_{\underset{\sim}{B}}(x)$	
•	代数积	algebraic product	$\mu_{\underset{\sim}{A}\cap\underset{\sim}{B}}(x)=\mu_{\underset{\sim}{A}}(x)\cdot\mu_{\underset{\sim}{B}}(x)=\mu_{\underset{\sim}{A}}(x)\mu_{\underset{\sim}{B}}(x)$	
$\oplus$	有界和	bounded sum	$a\oplus b=\min(a+b,1)$，式中 $a,b\in[0,1]$	$\gamma\geqslant0, p>0$
$\otimes$	有界积	bounded product	$a\otimes b=\max(a+b-1,0)$，式中 $a,b\in[0,1]$	$\gamma\geqslant0, p>0$
$\overset{+}{\varepsilon}$	爱因斯坦和	Einstein's sum	$a\overset{+}{\varepsilon}b=\dfrac{ab}{1+ab}$	式中 $a,b\in[0,1]$，$\gamma\geqslant0, p>0$
$\dot{\varepsilon}$	爱因斯坦积	Einstein's product	$a\dot{\varepsilon}b=\dfrac{ab}{1+(1-a)(1-b)}$	式中 $a,b\in[0,1]$，$\gamma\geqslant0, p>0$

符号	中文名称	英文名称	意义或举例	备注
$\overset{+}{\gamma}$	伽玛和	gamma sum	$a\overset{+}{\gamma}b=\dfrac{a\overset{\wedge}{+}b-(1-\gamma)ab}{\gamma-(1-\gamma)(1-ab)}$	式中 $a,b\in[0,1]$，$\gamma\geqslant0,p>0$
$\dot{\gamma}$	伽玛积	gamma product	$a\dot{\gamma}b=\dfrac{ab}{\gamma+(1-\gamma)(a\overset{\wedge}{+}b)}$	式中 $a,b\in[0,1]$，$\gamma\geqslant0,p>0$
$\overset{Y}{p}$	雅格和	Yager sum	$a\overset{Y}{p}b=\min(1,(a^p+b^p)^{1/p})$	式中 $a,b\in[0,1]$，$\gamma\geqslant0,p>0$
$\overset{\lambda}{p}$	雅格积	Yager product	$a\overset{\lambda}{p}b=1-\min(1,((1-a)^p+(1-b)^p)^{1/p})$	式中 $a,b\in[0,1]$，$\gamma\geqslant0,p>0$
$\sqcup$	取大运算	operation of fetch large	$\underset{\sim}{m}\sqcup\underset{\sim}{n}=\int_R\mu_{\underset{\sim}{m}}(x)\wedge\mu_{\underset{\sim}{n}}(y)/x\vee y$，$\underset{\sim}{m}$，$\underset{\sim}{n}$ 分别表示模糊数，即 $\underset{\sim}{m}=\int_R\mu_{\underset{\sim}{m}}(x)/x$，$\underset{\sim}{n}=\int_R\mu_{\underset{\sim}{n}}(x)/y$	
$\sqcap$	取小运算	operation of fetch small	$\underset{\sim}{m}\sqcap\underset{\sim}{n}=\int_R\mu_{\underset{\sim}{m}}(x)\vee\mu_{\underset{\sim}{n}}(y)/x\wedge y$	
$\neg$	减法运算	operation of subtraction	$\neg\underset{\sim}{m}=\int_R\mu_{\underset{\sim}{m}}(x)/(1-x)$	
$\rightsquigarrow$	模糊映射	fuzzy mapping	$f:X\rightsquigarrow Y$ 表示从 X 到 Y 的模糊函数	不同的场合中，模糊函数常有不同的定义
$\ominus$	有界差	bounded difference	$(A\ominus B)(x)=\max\{0,A(x)-B(x)\}$	
$\underset{\sim}{\leqslant}$	小于等于的放宽	relax restrictions of less or equal	$Ax\underset{\sim}{\leqslant}b\ (x\geqslant0)$ 表示约束条件 $Ax\leqslant b,x\geqslant0$ 的软化	
$D_{\rm fix}$	不动度	fixed degree	$D_{\rm fix}(x,F)=\alpha$，表示 x 关于模糊映射 $F:X\mathscr{F}w^{+}(X)$ 的不动度为 α，$\mathscr{F}(X)$ 表示 X 上所有模糊集组成的集	
e^*	绝对误差	absolute error	$e^*=x^*-x$，式中 x 表示精确值，x^* 为 x 的近似值	常简称误差
ε^*	误差限	limit of approximate value	$\lvert x^*\rvert<\varepsilon^*$，式中 x^* 为 x 的近似值，ε^* 为近似值 x^* 的误差限	
e_r^*	相对误差	relative error	$e_r^*=\dfrac{e^*}{x^*}$，式中 x 表示精确值，e^* 表示 x 的绝对误差，e_r^* 表示相对误差，它表示误差 e^* 关于近似值 x^* 的近似程度	
ε_r^*	相对误差限	limit of relative error	$\lvert e_r^*\rvert<\varepsilon_r^*$，式中 e_r^* 表示相对误差	
δ	最大相对误差	maximal relation error	$\lvert e_r^*\rvert=\dfrac{\lvert e^*\rvert}{\lvert x^*\rvert}\leqslant\delta$，式中 x^* 表示近似值，e^* 和 e_r^* 分别表示绝对误差和相对误差，取不等式成立的最小数 δ 为最大相对误差	
σ	标准误差	standard error	$\sigma=\sqrt{\dfrac{\sum\limits_{i=1}^{n}(x_i-x)^2}{n}}$，式中 $\sum\limits_{i=1}^{n}(x_i-x)^2$ 为误差平方和	
η	平均误差	mean error	$\eta=\dfrac{\sum\limits_{i=1}^{n}\lvert x_i-\bar{x}\rvert}{n}$，式中 $\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$ 是算术平均值	
v_i	离差	dispersion	$v_i=x_i-\bar{x}\quad(i=1,2,\cdots,n)$	
ν	概率误差	probabilistic error	$P(\lvert a\rvert\leqslant\nu)=1/2$ 表示数 a 的绝对值大于它的误差和小于它的误差出现的可能性一样大	
PS	多项式组	polynomial set	PS 表示由有限个非零多项式构成的集合	
Zero(•)	多项式的公共零点集	zero points set of polynomials	Zero (PS)表示多项式组 PS 中的多项式的公共零点集	
Res	结式	resultant	$\mathrm{Res}(p,q,x)=a_n^k b_k^n\cdot\prod\limits_{i=1}^{n}\prod\limits_{j=1}^{k}(\alpha_i-\beta_j)$．式中 α_i,β_j 分别是多项式 $p(x)$ 和 $q(x)$ 的根，$a_1,a_2,\cdots,a_n$ 和 $b_1,b_2,\cdots,b_k$ 分别为 $p(x)$ 和 $q(x)$ 的系数	

数 学 符 号 表

符号	中文名称	英文名称	意义或举例	备注
$\bar{\cup}$	合一运算	unification	$a\bar{\cup}b=a$,式中 a,b 均为原子,当且仅当 $a=b$ 时成立,否则 $a\bar{\cup}b$ 为空,集合论中的并运算是合一运算的特殊情况.	当原子不可分解时,合一的结果等于并集
R	羡余度	redundancy	$R=1-\frac{H_\infty}{H_0}$,式中 R 表示语言的羡余度,H_∞是极限熵,H_0 是语言成分等概率不相关时的熵	亦称冗余度
$E_t^{(p)}$	p 次指数平滑值	exponential smoothing value of pth	$E_t^{(1)} = \alpha\sum(1-\alpha)^i E_{t-i}^{p-1}$ $(p=2,3,\cdots)$,其中 $\alpha(1-\alpha)^i$ $(i=0,1,2,\cdots)$为当期序列值的影响权数,α 的一般范围在区间[0.1,0.5]内,适当选取 α 的值是保证预测的关键	当 $p=1$ 时即为一次指数平滑值 $E_t^{(1)}$
$\omega_t^{(p)}$	p 次加权平滑值	weight smoothing value of pth	$w_t^{(p)} = \alpha_0\sum_{i=1}^{\infty} d_i w_{t-i}^{p-1}$ $(t=\cdots,-1,0,1,\cdots,T)$,其中 $\alpha_i\,(i=0,1,2,\cdots)$为当期序列值的影响权数,$\alpha\in$[0.1,0.5]	当 $p=1$ 时为一次加权平滑值
VIF	协方差扩大因子	amplification factor of covariance	$\mathrm{VIF}(\beta_i)=\frac{1}{1-R^2}$,式中 β_i 为线性回归模型 $y=X\beta+\varepsilon$ 中 X 的第 i 个消费者预算参数 β_i 的估计值,R 为 X 的多重相关系数	
$r_u(x)$	风险厌恶度量	risk aversion measure	$r_u(x)=-\frac{u''(x)}{u'(x)}$,式中 u 为消费者的效用函数,自变量 x 可理解为收入	亦称 Arrow-Pratt 风险厌恶度量
S	价格单纯形	price simplex	$S=\{\boldsymbol{p}\in \mathrm{R}^l \mid p_k\geqslant 0,\sum_{k=1}^{l}p_k=1\}$,式中 R^l 是商品空间,$\boldsymbol{p}$ 表示价格向量	
β_i	预算映射	budget mapping	$\beta_i(p)=\{x\in X_i \mid p\cdot x\leqslant p\cdot e_i+\sum_{j=1}^{n}\theta_{ij}\pi_j(p)\}$,式中 $\beta_i(p)$和 X_i 分别表示第 i 个消费者的预算映射和消费集,π_j 是第 j 个生产者的利润函数	
a_{ij}	直接消耗系数	direct consumption coefficient	$a_{ij}=\frac{x_{ij}}{x_j}$ $(i,j=1,2,\cdots,n)$,x_{ij}表示第 i,j 两个部门的流量,x_j 表示第 j 个部门的总产品量	
b_{ij}	完全消耗系数	total consumption coefficient	$b_{ij}=a_{ij}+\sum_{k=1}^{n}a_{ik}a_{kj}+\sum_{s=1}^{n}\sum_{k=1}^{n}a_{is}a_{sk}a_{kj}+\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{k=1}^{n}a_{it}a_{ts}a_{sk}a_{kj}+\cdots(i,j=1,2,\cdots,n)$,式中 a_{ij}是直接消耗系数,b_{ij}表示第 j 个产品部门对第 i 种产品的完全消耗系数	
c_{ij}	完全需求系数	total demand coefficient	$c_{ii}=1+b_{ij}c_{ij}-b_{ij}(i\neq j)$,表示产品部门提供单位最终产品对所有产品部门产品的需求量,b_{ij}表示第 i,j 两个产品部门之间的完全消耗系数,c_{ij}表示第 j 个产品部门产出单位最终产品对第 i 个产品部门的需求量	
d_{ij}	投资系数	investment coefficient	动态投入产出模型中常用的统计指标,$d_{ij}=\frac{k_{ij}^t}{x_j^{t+1}-x_j^t}$,表示在 $t+1$ 时第 $j(j=1,2,\cdots,n)$部门增加单位产品需要第 i 投资部门在时间 t 供给第 j 部门产品的数量. k_{ij}^t表示 t 时 i 投资部门供给 j 部门产品总量,x_j^t 表示 j 部门 t 时的产品总量	
$L_{项}$	时滞	time lag	$L_{项}=[\alpha_1(n-0.5)+\alpha_2(n-1.5)+\cdots+\alpha_n 0.5]/100$ 为项目投资时滞,其中 α_i 为第 i 年投资占总投资的比重,n 为建设周期	
$L_{年}$	时滞	time lag	$L_{年}=\sum_{i=1}^{n}I_i n_i\Big/\sum_{i=1}^{n}I_i$ 为全年总投资时滞,式中 I_i 分配到 i 部门的投资. n_i 为 i 部门以外为单位的时滞	
$\boldsymbol{\varepsilon}$	应变张量	strain tensor	$\varepsilon_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)$ $(i,j=1,2,3)$,x_i,x_j 表示应变张量分量,u_i,u_j 表示位移分量	
k	高斯常数	Gauss constant	k≈0.017 202 098 95	
$\triangleq$	专用等号	symbol for special use	$a\oplus b\triangleq\max\{a,b\}$;$a\otimes b\triangleq a+b$ 表示极大代数中加法和乘法的定义	

符号	中文名称	英文名称	意义或举例	备注
Tayl	尾部	tail	$f=J^kf+\mathrm{Tayl}\,f$,式中 J^kf 是 f 在原点的泰勒展开式中保留 k 阶以下的多项式部分,截去的部分称为 f 的尾部,记为 Tayl f	
#()	袋	bag	$\#(x,B)$表示元素 x 在袋 B 中出现的次数.$\forall x\in B$,$0\leqslant\#(x,B)\leqslant1$ 时,袋 B 就蜕化为普通集合 B	
$W(s)$	传递函数	transfer function	$W(s)=\frac{Y(s)}{U(s)}=\frac{Q(s)}{P(s)}$,式中 $Y(s)$,$U(s)$分别为输出量和输入量的拉普拉斯变换式,$Q(s)$,$P(s)$分别为$W(s)$的分子、分母多项式	
cond	条件数	condition number	称 cond $G=\frac{\sigma_{\max}}{\sigma_{\min}}\geqslant1$ 为矩阵 G 的条件数,cond G越大,矩阵 G 越趋于欠秩	
diag	对角元	diagonal element	设 $S=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_p)$,则称 $\sigma_i(i=1,2,\cdots,p)$为对角矩阵 S 的对角元	
blockdiag	块对角元	block diagonal element	设 $X=$ block diag $(\Delta_1,\cdots,\Delta_1,\Delta_2,\cdots,\Delta_2,\cdots,\Delta_r,\cdots,\Delta_r)$,其中 Δ_i 为 k_i 阶方阵,则称 Δ_i 为块对角矩阵的块对角元	
arg(・)	相角	phase angle	$\arg(g(\mathrm{j}\omega))$称为相角,其中$g(\mathrm{j}\omega)$为 $m\times n$ 阶复阵函数,j 为虚数单位	
conv(・)	凸包	convex hull	conv $f(\mathrm{j}\omega,\Gamma)=$conv $f(\mathrm{j}\omega,\Gamma_0)$,式中 conv(・)表示$\mathrm{R}^2$ 上的凸包,$\omega\in\mathrm{R}$,j 为虚数单位,$\Gamma_0\triangleq\{\nu\mid\nu_i=0,1;\ i=1,2,\cdots,m\}$为 $\nu_i(i=1,2,\cdots,m)$中的多仿射函数	
$\asymp$	等序关系	equals order relation	若 z_1,z_2 为两个非零复数,且$\frac{z_2}{z_1}\neq0$,则记为 $z_1\asymp z_2$	
ess sup	本质上确界	essential supremum	$\mathrm{ess}\sup\limits_{\omega}\sigma(G(\mathrm{j}\omega))$表示 $m\times n$ 阶复矩阵值函数$G(\mathrm{j}\omega)$的本质上确界,即除去 ω 的一个零测子集后的上确界	
s.t	约束条件	constraint condition	$\max f=\sum\limits_{j=1}^{n}c_jx_j$, $\text{s.t}\begin{cases}\sum\limits_{j=1}^{n}a_{ij}x_j=b_i & (i=1,2,\cdots,m),\\ x_j\geqslant0 & (j=1,2,\cdots,n),\end{cases}\quad(*)$. 目标函数 max f 必须满足(*)中的条件	
$\overset{L}{>}$	字典序	lexicographicall order	$V\overset{L}{>}0$ 表示字典式为正的;$V\overset{L}{<}0$ 表示字典式为负的;Lex min 表示字典式最小	$V=(v_1,v_2,\cdots,v_n)$是 n 维向量空间的向量
$\underset{\smile}{\delta}_B$	下特征数	low characteristic number	$\underset{\smile}{\delta}_B=\begin{cases}\max\left\{-\frac{\lambda_i}{\lambda_j^*}\middle\|\lambda_j^*<0\right\} & (\exists\ \lambda_j^*<0),\\ -\infty & (\bar{\exists}\ \lambda_j^*<0),\end{cases}$ $\underset{\smile}{\delta}_B$ 称为基 B 的下特征数 λ_j,λ_j^* 为检验数	
$\overset{\frown}{\delta}_B$	上特征数	above characteristic number	$\overset{\frown}{\delta}_B=\begin{cases}\min\left\{-\frac{\lambda_i}{\lambda_j^*}\middle\|\lambda_j^*>0\right\}, & \exists\ \lambda_j^*>0,\\ +\infty, & \bar{\exists}\ \lambda_j^*>0,\end{cases}$ $\overset{\frown}{\delta}_B$ 称为基 B 的上特征数,λ_i,λ_j^* 为检验数	
⋙	等级标志关系	relation of order mark	$p_i⋙p_j$ 表示在一个单目标函数 min $f=p_1f_1+p_2f_2+\cdots+p_lf_l$ 中,$p_1,p_2,\cdots,p_l$ 为等级标志关系	
$P(\cdot)$	策略	policy	P 表示最优策略.$P_{k,n}^*(x_k)$表示最优子策略,是初始状态为 x_k 的后部子过程所有子策略中最优者	
opt	最优值	optimum value	opt $v_{k,n}[x_k,P_{k,n}(x_k)]$表示指标函数 $v_{k,n}$的最优值,$P_{k,n}$表示子策略是从第 k 段开始到终点过程的策略	
pos	正线性组合集	set of positive linear combination	$\mathrm{pos}A=\{\alpha\mid\alpha\in\mathrm{R}^m,\ \alpha=\sum\limits_{j=1}^{n}\beta_jA_j,\ \beta_j\geqslant0,j=1,2,\cdots,n\}$ 表示由矩阵 A 的各列的正线性组合组成的集合	
epi	上图	epigraph	epi $f=\{(x,\alpha)\mid\alpha\geqslant f(x)\}$表示函数 $f(x)(x\in\mathrm{R}^n)$的上图,若给定 epi f,则 $f(x)=\min\{x\mid(x,\alpha)\in\mathrm{epi}\ f\}$	
/	排队记法	queueing notation	$X/Y/Z/C$ 为排队记法,其中 X,Y,Z,C 的意义依次为:1. 相继到达间隔时间的分布;2. 服务时间的分布;3. 服务台的数目;4. 允许的顾客容量	

符号	中文名称	英文名称	意义或举例	备注
Ls	队长期望值	team length expected value	$Ls=\sum_{n=0}^{\infty}nP_n=\frac{\rho}{1-\rho}=\frac{\lambda}{\mu-\lambda}$表示标准的$M/M/1$模型的队长期望值，$\rho$为服务强度，即服务台平均利用率	
Lq	队列长期望值	queueing length expected value	$Lq=\sum_{n=1}^{\infty}(n-1)P_n=Ls-P=\frac{\rho^2}{1-\rho}=\frac{\rho\lambda}{\mu-\lambda}$表示标准的$M/M/1$模型的队列长期望值，$\rho$为服务强度，即服务台平均利用率	
Ws	逗留时间期望值	expected value of staying time	$Ws=E[W]=\frac{1}{\mu-\lambda}$表示标准的$M/M/1$模型的逗留时间期望值	
Wq	等待时间期望值	expected value of waiting time	$Wq=Ws-\frac{1}{\mu}=\frac{\rho}{\mu-\lambda}$表示标准的$M/M/1$模型的等待时间期望值	
G	对策	games	对策$G=(S_1,S_2,A)$，其中$S_1=\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$表示局中人Ⅰ的纯策略集合，$S_2=\{\beta_1,\beta_2,\cdots,\beta_n\}$表示局中人Ⅱ的纯策略集合. $A=(a_{ij})_{m\times n}$表示支付(赢得)矩阵	
V_G	对策值	games value	$V_G=\max_i\min_j a_{ij}=\min_j\max_i a_{ij}$称为对策$G=\{S_1,S_2,A\}$的值	
Te	噪声温度	noise temperature	$Te=\frac{N}{kB}(k)$，其中N为噪声功率，k为玻耳兹曼常数，B为频带宽度(Hz)	
γ	传播常数	propagation constant	$\gamma=\alpha+j\beta=\sqrt{Z_1r_1}$，其中$\alpha$表示衰减常数(Np/m，dB/M)，$\beta$表示相移常数(rad/m)	
Lt	传输损耗	loss of transmission	$Lt=32.45+20\lg f+20\lg d+A-G_t-G_r$，式中$f$为工作频率(MHz)，$d$为传输距离(km)，$A$为电路衰减(dB)，$G_t$、$G_r$分别为发射天线与接收天线的增益(dB)	
C	信道容量	channel capacity	$C=\max_{P(x)}I(x;y)$，其中$P(x)$为输入符号概率(或概率密度)，$I(x;y)$为互信息量	
$R(D^*)$	信源率失真函数	source rate distortional function	$R(D^*)=\min\{I(u;v)\}$，$P(v_j\|u_j)\in B_D$，其中D^*为信源的允许平均失真度，$I(u;v)$为平均互信息量	
I_A	自信息量	self-information	$I_A=\log\frac{1}{P(A)}=-\log P(A)$，式中$P(A)$为随机事件$A$发生的概率，$I_A$表示$A$的自信息量	
$I(x;y)$	互信息量	mutual information	$I(x;y)=\log\frac{P(x\|y)}{P(y)}$，式中$y$表示收到的消息，$x$表示收到消息的某事件的信息量	
$I(X;Y)$	平均互信息量	average mutual information	$I(X;Y)=H(X)-H(X\|Y)$，其中$H(X)$代表接收到输出符号集Y以前关于输入符号集X的平均不确定性；$H(X\|Y)$代表接收到输出符号集Y后关于输入符号集X的平均不确定性	
$\oplus$	逻辑导式运算符号	operational symbol of logical derived rule	$D(\alpha,\beta)=\sum_{i=1}^{n}(\alpha_i\oplus\beta_i)$，式中$\alpha_i$，$\beta_j$表示长度为$n$的二进制序列码元，$\alpha_i\oplus\beta_i$是二进制码元相加. $D(\alpha,\beta)$表示α，β对应位置上码元取值不同的个数	
$\circledast$	周期卷积	periodic convolution	$\tilde{x}_1(n)\circledast\tilde{x}_2(n)$，式中$\tilde{x}_1(n)$和$\tilde{x}_2(n)$表示周期长度	
Ⓝ	循环卷积	circular convolution	$\tilde{x}_1(n)$Ⓝ$\tilde{x}_2(n)$	

撰 稿 王怀安 刘宝康 杨子胥 杨德平
段 方 郝拉娣 阎崇正
审 定 李志深 陈惠津 阎崇正

条目笔画索引

说明：1. 该索引收录了本卷正文中给出释文的全部条目及其参见条目，提供读者按汉字笔画方式检索使用。

2. 以汉字起首的条目标题按第一字的笔画由少到多的顺序排列，若笔画数相同，则按一（横）、丨（竖）、丿（撇）、丶（点）、乛（折）五种笔形顺序排列，其中，㇀（提）归为一（横），亅（竖钩）归为丨（竖），㇏（捺）归为丶（点），各种笔形带钩或曲折的笔画（除竖钩"亅"外）归为乛（折）。第一个字相同的，则按第二个字的笔画数和起笔笔形的顺序排列，依次类推。

3. 凡第一个字为西文字母、数学符号、罗马数字和阿拉伯数字起首的条目标题，一律排在汉字起首条目标题的最后。以西文字母起首的条目标题分别按其字母的花体、大写、小写及字母本身的先后顺序排列；数学符号起首的条目标题按知识结构顺序排列；数字起首的条目标题按由小到大的顺序排列。若起首的字母、符号及数字相同时，仍按其后汉字的笔画顺序排列。

一　画

二　画

三 画

四　画

五　画

六 画

七 画

八　画

九　　画

十　画

十　画

十 一 画

十 二 画

十　三　画

十 四 画

十 五 画

十 六 画

十七画

十八画

十九画

二十画

其他

条目音序索引

说明：1. 该索引收录了本卷正文中给出释文的全部条目及其参见条目，提供读者按汉语拼音方式检索使用。

2. 以汉字起首的条目标题按第一字的汉语拼音字母顺序排列，若第一字的声母、韵母相同，则按声调的阴平、阳平、上声、去声顺序排列。第一个字相同的，则按第二个字的汉语拼音字母顺序排列，多音字按不同的拼音字母顺序排列，依此类推。

3. 凡第一个字为西文字母、数学符号、罗马数字和阿拉伯数字起首的条目标题，一律排在汉字起首条目标题的最后。以西文字母起首的条目标题分别按其字母的花体、大写、小写及字母本身的先后顺序排列；数学符号起首的条目标题按知识结构顺序排列；数字起首的条目标题按由小到大的顺序排列。若起首的字母、符号及数字相同时，仍按其后汉字的拼音字母顺序排列。

A

B

C

D

E

F

G

H

J

K

L

M

N

O

P

Q

R

T

W

X

Y

Z

其　　他

条目西文索引

说明：1. 该索引收录了本卷正文中给出西文标题的全部条目，提供读者按西文检索使用.

2. 条目标题按起首西文字母的顺序排列(同一字母先大写，后小写)；条目标题的西文缩写，按一个词排列。其他文种亦按此原则编排。

3. 凡以数学符号、罗马数字和阿拉伯数字起首的条目标题，一律排在条目西文索引的最后。数学符号起首的条目标题按知识结构顺序排列；数字起首的条目标题按由小到大的顺序排列。

4. 若条目标题起首的字母、符号、数字相同时，则按第二个字母等的顺序排列，余此类推。

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

W

X

Y

Z

其　　他

中外人名译名对照表

A

阿波罗尼奥斯(Apollonius,(P))
阿德尔曼(Adleman)
阿德勒(Adler,R. L.)
阿尔文(Alfvén,H.)
阿格兰德(Argelander,F. W. A.)
阿基米德(Archimedes)
阿克曼(Ackermann,F.)
阿克曼(Ackermann,J.)
阿克曼(Ackermann,W.)
阿克赛罗德(Axelrod,R.)
阿拉果(Arago,D. F.)
阿罗(Arrow,K,J.)
阿罗(Haro,G.)
阿姆巴楚米扬(Амбарцумян,В. А.)
阿诺尔德(Арнольд,В. И.)
阿什贝(Ashby,W. R.)
阿斯兰卡斯克维里(Aslanikaskvili)
阿依热尔曼(Айзерман,М. А.)
埃尔朗(Erlang,A. K.)
埃根松(Эйгенсон,М. С.)
埃克特(Eckert,J. P.)
埃奇沃思(Edgeworth,F. Y.)
埃萨克斯(Isaacs,R.)
埃文斯(Evans,W. R.)
艾根(Eigen,M.)
艾里(Airy,G. B.)
艾利斯伯格(Ellisberg)
艾萨克斯(Issacs,R.)
爱丁顿(Eddington,A. S.)
爱因斯坦(Einstein,A.)
安德罗诺夫(Андронов,А. А.)
昂萨格(Onsager,L.)
奥本海默(Oppenheimer,J. R.)
奥伯斯(Olbers,H. W. M.)
奥尔特(Oort,J. H.)
奥斯特隆姆(Astrom,K. J.)
奥文斯(Owens,A. J.)

B

巴泊柯维奇(Папкович,П. Ф.)
巴德(Baade,W. H. W.)
巴尔达(Baarda,W.)
巴科斯(Backus,J.)
巴拉修斯(Palacios,F. G.)
巴米希(Barmish,B. R.)
巴纳德(Barnard,E.)
巴日诺夫(Важнов,А.)
巴斯科特(Baskett,F.)
巴特切亚(Bhattacharyya,S. P.)
柏拉图(Plato)
柏萝登(Broyden,C. G.)
拜利(Bailey,F. N.)
班勒卫(Painlevé,P.)
邦迪(Bondi,H.)
鲍达(Borda,J. G.)
鲍恩(Bowen,I. S.)
鲍威尔(Powell,M. J. D.)
鲍文(Bowen,R.)
贝蒂(Betti,E.)
贝尔(Beale,E. M. L.)
贝尔(Bell,A. G.)
贝尔(Bell,D. E.)
贝尔(Bell,J. L.)
贝尔曼(Bellman,R.)
贝尔特拉米(Beltrami,E.)
贝克(Bak,P.)
贝克(Beke,E. V.)
贝克林(Becklin,E. E.)
贝雷模(Bereamu,B.)
贝拿扬(Benayoun,R.)
贝纳德(Benard,C.)
贝塞尔(Bessel,F. W.)
贝特朗(Bertrand,G.)
本迪克森(Bendixen,A.)
本斯(Byrnes,C.)
本温尼斯特(Benveniste,A.)
比维斯(Bevis,J.)
彼得森(Peterson,E.)
波波夫(Попов,А. С.)
波波夫(Попов,В. М.)
波波特(Rapoport,A.)
波得(Bode,J. E.)
波加提(Burgatti)
波拉克(Pawlak,Z.)
波莱尔(Borel,(F.-É.-J.-)É.)
波利亚(Polya,G.)
波林(Bohlin,K.)
波斯特(Post,E. L.)
波斯特尼考夫(Постников,В. Н.)
波特(Bode,H. W.)

中外人名译名对照表

玻耳兹曼(Boltzmann,L. E.)
伯比奇夫妇(Burbidge,G. and Burbidge,E. M.)
伯德(Bode,H. W.)
伯格森(Bergson,A.)
伯克霍夫(Birkhoff,G. D.)
伯姆(Böhm,C.)
伯斯姆特(Bismut)
泊松(Poisson,S. -D.)
勃龙别格(Брумберг,В. А.)
勃罗森(Brorsen,T.)
博杜恩·德·库尔特内(Baudouin de Courtenay,J. N.)
博雷尔(Borel,C.)
博斯(Boss,B.)
博斯(Boss,L.)
布尔(Boole,G.)
布丰(Buffon,G. -L. L.)
布拉德雷(Bradley,J.)
布拉萨德(Brassard)
布兰德(Bland,R. G.)
布朗(Brown,E.)
布劳威尔(Brouwer,D.)
布列斯南(Bresnan,J.)
布龙菲尔德(Bloomfield,L.)
布伦斯(Bruns,H.)
布罗贝(Broadbent,S. R.)
布洛根(Brogan,W. L.)
布尼亚科夫斯基(Буняковский,В. Я.)
布森内斯克(Boussinesq,J. V.)
布特柯夫斯基(Butkovskiy,A. G.)
布西(Bucy,R. S.)

C

策尔夫(Cerf,V.)
策梅洛(Zermelo,E. F. F.)
查恩斯(Charnes,A.)
查普利特(Chapellat,H)
查森(Richardson,L. F.)
柴培耳(Zeipel,H. von.)
昌德拉塞卡(Chandrasekhar,S.)
陈(Chen,P. P. S.)
陈鹤琴(Chen Heqin)
赤池弘次(Akaike)

D

达尔文(Darwin,G. H.)
达芬(Duffin,R. J.)
达朗贝尔(d'Alembert,J. le R.)
达维靖可夫(Давиденков,Н. Н.)
戴肯(Daikin,R. J.)
戴维顿(Davidon,W. C.)
戴维森(Davison,E. J.)
戴维士(Davis,L.)
戴文赛(Dai Wensai)
丹尼尔第一·伯努利(Bernoulli,Daniel Ⅰ)
丹齐克(Dantzig,G. B.)
丹戎(Danjon,A.)
得叶(Dyer,I. S.)
德·菲莫蒂(De Fimetti,B.)
德·摩根(De Morgen,A.)
德布鲁因(de Bruijn,N. G.)
德布罗(Debreu,G.)
德拉帕(Draper,C. S.)
德雷耶(Dreyer,J. L. E.)
德洛内(Delaunay,C. E.)
德沃库勒(de Vaucouleurs,G.)
邓聚龙(Deng Julong)
狄克(Dicke,R. H.)
狄喇克(Dirac,P. A. M.)
迪丁贝尔格(Dittinberger,W.)
迪费(Diffie,W.)
迪克斯特拉(Dijkstra,E. W.)
笛卡儿(Descartes,R.)
帝克斯彻(Dijkstra,E. W.)
第谷(Tycho Brahe)
丁伯根(Tinbergen,J.)
丁渭(Ding Wei)
多夫曼(Dorfman,R.)
多伊尔(Doyle,J. C.)

E

额舍耳比(Eshelby,J. D.)
厄得蒙斯(Edmonds,J.)
厄莫里叶夫(Ermoliev,Y.)
厄农(Henon,M.)
恩盖塞(Engesser,F.)
恩克(Encke,J. F.)

F

伐里尼翁(Varignon,P.)
法布里修斯(Fabricius,J.)
法拉第(Faraday,M.)
范·德·赫斯特(Van de Hulst,H. C.)
范艾伦(Van Allen)
范登贝尔格(Van den Bergh,S.)
范莱因(Van Rhijn,P. J.)
菲尔德(Field,R. J.)
菲尔墨(Fillmore,C. J.)
斐波那契(Fibonacci,L.)
费尔得鲍姆(Feldbaum,A. A.)
费根鲍姆(Feigenbaum,E. A.)
费根鲍姆(Feigenbaum,M. J.)
费马(Fermat,P. de)
费米(Fermi,E.)
费史伯恩(Fishburn,P. C.)
冯·贝塔朗菲(von Bertalanffy,L.)
冯·诺伊曼(von Neumann,J.)
冯·斯塔克伯格(Von Stackelberg,H.)
冯·西利格(Seeliger,H. von)
冯康(Feng Kang)
佛科(Foucault,J. L.)

弗格尔(Fogel,L. J.)
弗凯拉(Fichera,G.)
弗莱彻(Fletcher,R.)
弗莱明(Fleming,W. P.)
弗兰克(Frank,P. H.)
弗兰克兰德(Frankland,E. F.)
弗朗希斯(Francis,B. A.)
弗里德里希斯(Friedrichs,K. O.)
弗里德曼(Фридман,A.)
弗里克(Fricke,W.)
弗里斯(Fliess,M.)
弗里特曼(Фридман,Я. В.)
弗洛伊德(Freud,G.)
弗斯特勒尔(Förstner,W.)
福勒(Fowler,W. A.)
福雷斯特(Forrester,J. W.)
福特(Ford,L. R.)
富尔克森(Fulkerson,D. R.)
富克斯(Fuchs,H. M. L.)
富纳西尼(Fornasini)

G

伽利略(Galilei,G.)
伽辽金(Галёркин,Б. Г.)
伽莫夫(Gamov,G.)
盖斯特(Gest,J. J.)
盖兹达(Gazdar,G.)
高斯(Gauss,C. F.)
戈德福布(Goldfarb,D.)
戈尔德(Gold,T.)
戈尔德伯革(Goldberg,D.)
哥白尼(Kopernik,M.)
哥德尔(Gödel,K.)
哥德曼(Goldman,A. J.)
哥拉沃(Glover,K.)
格拉法伦(Grafarend,E. W.)
格莱比(Graebe,S. F.)
格莱斯堡(Gleissberg,W.)
格朗沃尔(Glover,K.)
格里菲思(Griffiths,A. A.)
格里芬斯塔特(Grefenstette,J.)
格里哥里十三世(Gregorius,XIII)
格林(Green,G.)
格莫理(Gomory,R. E.)
古德里克(Goodricke,J.)
古斯(Guth,A. H.)
郭守敬(Guo Shoujing)

H

哈勃(Hubble,E. P.)
哈德利(Hadley,J.)
哈丁(Harding,C. L.)
哈格(Hagg,G.)
哈根斯(Huggins,S.)
哈肯(Haken,H.)
哈雷(Halley,E.)
哈里斯(Harris,F.)
哈里托诺夫(Харитонов,В. Л.)
哈马逊(Humason,M.)
哈密顿(Hamilton,W. R.)
哈默斯利(Hammersley,J. M.)
哈姆(Halm,J. K. E.)
哈奇扬(Хачиян,Л. Г.)
哈特利(Hartley,R. V. I.)
海拉瓦(Helava,U.)
海利-肖(Hele-Shaw,H. S.)
海姆斯瓦尔斯(Himsworth,F. R.)
海普尔(Hipel,K. W.)
亥姆霍兹(Helmholtz,H. L. F.)
汉夫金斯(Hawkins,D. J.)
汉默(Hammer,J.)
汉森(Hansen,P. A.)
豪克斯(Hawks,J.)
豪斯多夫(Hausdorff,F.)
何毓琦(He Yuqi)
贺布兰德(Herbrand,J.)
赫比格(Herbig,G. H.)
赫伯(Hebb,D.)
赫茨普龙(Hertzsprung,E.)
赫尔(Hoare,C. A. R.)
赫尔德(Hölder,O. L.)
赫尔曼(Hellman,M. E.)
赫尔默特(Helmert,F. R.)
赫尔维茨(Hurwitz,A.)
赫林格(Hellinger,E.)
赫瑞鲍(Horrebow,C.)
赫斯蒂(Hext,G. R.)
赫斯泰尼斯(Hestenes,M. R.)
赫维茨(Hurwicz,L.)
赫歇耳(Herschel,J. F. W.)
赫兹(Hertz,H. R.)
赫兹普龙(Hertzsprung,E.)
黑尔(Hale,G. E.)
黑塞(Hesse,L. O.)
亨德森(Henderson,T.)
侯伯特(Houbolt,J. C.)
胡伯(Huber,M. T.)
胡海昌(Hu Haichang)
胡克(Hooke,R.)
华尔特(Wilde,D. J.)
华罗庚(Hua Loo-Keng)
黄授书(Huang Shoushu)
惠更斯(Huygens,C.)
惠特尼(Whitney,H.)
霍尔(Hall,A. D.)
霍尔(Hall,C. A. R.)
霍夫(Hoff,M. E. Jr.)
霍夫曼(Hoffman,A. J.)
霍华德(Howard,R. A.)
霍兰(Holland,J.)

霍兰德(Holland,J.)
霍普菲尔德(Hopfield,J. J.)
霍普夫(Hopf,E.)
霍伊尔(Hoyle,F.)

J

基尔霍夫(Kirchhoff,G. R.)
基夫斯(Jeeves,T. A.)
基弗(Kiefer,J. C.)
基罗(Guiraud,P.)
吉布斯(Gibbs,J. W.)
加勒(Galle,J. G.)
加斯(Gass,S.)
贾耽(Jia Dan)
贾克比尼(Jacopini,G.)
杰夫里翁(Geoffrion,A. M.)
金斯(Jeans,J. H.)
爵德曼(Joan Daemen)

K

卡尔(Kall,P.)
卡尔曼(Kalman,R.)
卡尔曼(Kalman,R. E.)
卡捷尔尼可夫(Котельников,В. А.)
卡灵顿(Carrington,R. C.)
卡马卡(Karmarkar,N.)
卡门(Kármán,T. von)
卡诺(Carnot,N. L. S.)
卡普兰(Kaplan,R.)
卡普坦(Kapteyn,J. C.)
卡斯提利亚诺(Castigliano,A.)
卡西尼(Cassini,J. D.)
开尔文(Kelvin,L.)
开普勒(Kepler,J.)
凯定(Kaeding,F. W.)
凯珀(Kuiper,G. P.)
凯塞(Kaiser)
凯依(Kay,M.)
坎贝尔(Campbell,L.)
坎托罗维奇(Канторович,Л. В.)
康德(Kant,I.)
康尼斯内(Konecny,G.)
康宁汉-格林(Cuninghame-Green,R.)
康普顿(Compton,A. H.)
康托尔(Cantor,G. (F. P.))
康托尔(Cantor,G.)
康维(Conway,J.)
康熙(Kang Xi)
考莫瑞尔(Colmerauer,A.)
柯恩(Korn,A.)
柯恩海姆(Konheim,A. G.)
柯尔莫哥洛夫(Колмогоров,А. Н.)
柯弗(Kopff,A.)
柯含农(Kohonen,T.)
柯尼希(König,D.)
柯娃尔斯基(Kowalski,R. A.)
柯瓦列夫斯卡娅(Ковалевская,С. В.)
柯西(Cauchy,A. -L.)
柯伊塔(Koiter,W. T.)
科恩(Cohen,G.)
科克(Koch,K. R.)
科拉索夫(Колосов,Г. В.)
科里奥利(Coriolis,G. G. de)
科林斯(Collins,G. E.)
科帕尔(Kopal,Z.)
科威耳(Cowell,P. H.)
克尔(Kerr,J.)
克拉伯龙(Clapeyrom,B. P. E.)
克拉克(Clarck,A. G.)
克拉美(Cramer,G.)
克拉索夫斯基(Красовский,Н. Н.)
克莱罗(Clairaut,A. -C.)
克劳斯(Cross,H.)
克雷洛夫(Крылов,Н. С.)
克里斯蒂(Christy,J. W.)
克列因(Крейн,М. Г.)
克林(Kleene,S. C.)
克吕格(Krüger,I. H. L.)
克罗蒂(Crotti,F.)
克洛梅林(Crommelin,A. C.)
克尼(Keeny,R. L.)
肯德尔(Kendall,D. G.)
寇赫(Koch,H. Von)
库伯(Cooper,W. W.)
库恩(Kuhn,H. W.)
库克(Cook,S. A.)
库列克(Kureck)
库伦(Coulomb,C. A. de)
库普曼斯(Koopmans,T. C.)
库斯唐赫莫(Kustaanheimo,K.)

L

拉波波特(Rapoport,A.)
拉弗(Love,A. E. H.)
拉格朗日(Lagrange,J. -L.)
拉卡伊(LaCaille,N. L. de)
拉朗德(Lalande,J. J. F. de)
拉马奇(Ramadge,P.)
拉曼(Laman,C. V.)
拉梅(Lamé,G.)
拉姆齐(Ramsey,F. P.)
拉姆钱达尼(Ramchandani,C.)
拉匹塔斯(Lapidus,L.)
拉普拉斯(Laplace,P. -S.)
拉塞尔(Lassel,W.)
拉叶(Rayet,G. A. P.)
莱昂斯(Lions,J. -L.)
莱布尼茨(Leibniz,G. W.)
莱德维奇(Ledwich,G.)
莱顿(Leighton,R. B.)

赖柴尔(Résal,H.)
赖斯(Rie,J. R.)
兰彻斯特(Lanchester,F. W.)
兰丁(Landin,P. J)
兰顿(Langton,C. G.)
兰法(Raiffa,H.)
朗伯(Lambert,J. H.)
朗道(Ландау,Л.)
劳斯(Routh,E. J.)
勒姆克(Lemke,C. K.)
勒奈夫(Lernev,A. Y.)
勒威耶(Le Verrier,U. J. J.)
勒维特(Leavitt,H. S.)
雷登(Luyten,W. J.)
雷切恩伯格(Rechenberg,I.)
黎曼(Riemann,(G. F.)B.)
黎子良(Li Ziliang)
李德仁(Li Deren)
李诚(Li jie)
李氏(Lie,M. S.)
李斯廷(Listing,J. B.)
李天岩(Li Tianyan)
李亚普诺夫(Ляпунов,A. M.)
李耀滋(Li Yaozi)
里夫斯(Reevse,C. M.)
里奇(Ritchie,D. M.)
理查森(Richardson,L. F.)
利齐奥里(Riccioli,J. B.)
列昂节夫(Leontief,W.)
林德布拉德(Lindblad,B.)
林登迈尔(Lindenmayer,A.)
林家翘(Lin Chia-Chiao)
林纳斯(Linus,B. T.)
龙伯格(Luenberger,D. G.)
鲁宾斯坦(Rubinstein,R.)
鲁宾孙(Robinson,J. A.)
鲁里叶(Лурье,А. И.)
鲁梅哈特(Rumelhart,D.)
吕卡(Lucas,F. -É. -A.)
吕兹勒(Rössler,O.)
罗(Rowe,A. P.)
罗宾斯(Robbins,H.)
罗伯特·肖(Shaw,R.)
罗伯逊(Robertson,H. P.)
罗洪先(Luo Hongxian)
罗塞尔(Roesser)
罗森(Rosen,J. B.)
罗森布拉特(Rosenblatt,F.)
罗森布罗克(Rosenbrock,H. H.)
罗斯(Ross,F. E.)
罗斯(Rosse,W. P.)
罗斯兰德(Rosseland,S.)
罗斯勒(Rossler)
罗素(Russell,H.)
洛基尔(Lockyer,J. N.)
洛克菲勒(Rockafellar,R.)
洛伦兹(Lorenz,E. N.)
洛特卡(Lotka,A. J.)
洛希(Roché,E. A.)

M

马尔可夫(Марков,А. А.)
马赫(Mach,E.)
马卡良(Маркарян,Б. Е.)
马可尼(Marconi,G.)
马克劳林(Maclaurin,C.)
马里奥特(Mariotte,E.)
马舍西尼(Marchesini)
马斯孙(Matheson,J. E.)
马特斯(Matthes,K.)
迈恩(Manne,A. S.)
迈耶(Meyer,K. R.)
麦卡洛克(McCulloch,W. S.)
麦卡锡(McCarthy,J.)
麦克安德鲁(McAndrew,M. H.)
麦克卡洛克(Mcculloch)
麦克斯韦(Maxwell,J. C.)
麦克维支(Markowitz,H.)
麦哲伦(Magellan,F.)
曼达尼(Mamdani,E. H.)
曼德勃罗特(Mandelbrot,B. B.)
梅得(Mead,R.)
梅卡托(Mercator,N.)
梅里尔(Merrill,P. W.)
梅里斯(Mirrless,J.)
梅林(Merlin,P.)
梅塞奇(Message,P. J.)
梅西耶(Messier,C.)
门登豪尔(Mendenhall,T. C.)
门罗(Monro,S.)
蒙兹(Muntz,R. R.)
蒙日(Monge,G.)
蒙塔鸠(Montague,R.)
蒙塔纳里(Montanari,G.)
米尔恩(Milne,E. A.)
米尔纳(Milner,R.)
米哈尔斯基(Michalski,R. S.)
米赫林(Михлин,С. Г.)
米勒(Miller,R. E.)
米诺尔(Milnor,J.)
米沙诺维奇(Mesarovic,M. D.)
米谢阿(Michea,R.)
米泽斯(Mises,R. von)
密特尔(Mitter,S. K.)
密歇尔斯基(Мещерский,И. В.)
明斯基(Minsky,M.)
摩根(Morgan,B. S.)
摩根(Morgen,W. W.)
莫尔(Mohr,O.)
莫尔(Moore,J. B.)

莫尔顿(Moulton,F. R.)
莫尔斯(Morse,S.)
莫尔斯(Morse,M.)
莫夫强(Мовчан,А. А.)
莫根施特恩(Morgensten,O.)
莫勒(Moler,C.)
莫里森(Morrison,P.)
莫洛金斯基(Молоденский,М. С.)
莫洛伊(Molloy,M. K.)
莫奇利(Mauchly,J. W.)
莫瑞拉(Morera,G.)
莫泽(Moser,J. K.)
木下宙(Kinoshita)
穆斯赫利什维里(Мусхелишвили,Н. И.)

N

纳什(Nash,J. F.)
纳什(Nash,N.)
纳维(Navier,C. (-L. -M. -H.))
奈奎斯特(Nyquist,H.)
奈曼(Neyman,J.)
内密洛夫斯基(Nemirovskii,A. S.)
尼古拉第二・伯努利(Bernoulli,Nicolaus Ⅱ)
尼柯尔斯(Nichols,N. B.)
涅尔得(Nelder,J. A.)
牛顿(Newton,I.)
牛顿(Newton,J.)
纽科姆(Newcomb,S.)
纽曼(Neumann,M.)
诺德威特(Nordtvedt,K. J.)
诺特(Noether,(A.)E.)
诺伊格鲍尔(Neugebauer,G.)
诺伊斯(Noyes,R. M.)

O

欧几里得(Euclid)
欧拉(Euler,L.)
欧文(Irwin,G. R.)

P

帕雷托(Pareto,V.)
帕雷托(Pareto,L.)
帕斯卡(Pascal,B.)
派垂(Petri,C. A.)
潘索(Poinsot,L.)
庞加莱(Poincaré,(J. -)H.)
庞特里亚金(Понтрягин,Л. С.)
培根(Bacon,R.)
裴秀(Pei Xiu)
佩瑞拉(Pereira,F.)
彭罗塞(Penrose,R.)
彭齐亚斯(Penzias,A.)
彭赛列(Poncelet,J. -V.)
皮尔逊(Pearson,K.)
皮特斯(Pitts,W.)
皮特逊(Peterson,J. L.)
皮特逊(Peterson,L.)
皮亚齐(Piazzi,G.)
匹特里(Petri,C. A.)
匹兹(Pitts,W.)
珀塞尔(Purcell,E. M.)
普朗托(Prandtl,L.)
普里考帕(Prekopa,A.)
普利高津(Prigogine,I.)
普森(Pogson,N. R.)
普希金(Pushkin,A.)

Q

齐奥尔柯夫斯基(Циолковский,К. Э.)
齐普夫(Zipf,G. K.)
钱学森(Qian Xuesen)
乾隆(Qian Long)
强迪(Chandy,K. M.)
乔姆斯基(Chomsky,N.)
乔万尼(Givone)
乔伊(Joy,A. H.)
切博塔廖夫(Чеботарёв,Г. А.)
切仑柯夫(Черенков,Л. А.)
钦斯特拉(Tienstra,J. M.)
琼(Jong,K. de)
丘宏义(Qiu Hongyi)
丘奇(Church,A.)
屈斯特纳(Küstner,K. F.)

R

茹勒(Ruelle,D.)
儒略(Julius)
儒略・恺撒(Caesar,G. J.)
瑞费斯特(Rivest)
瑞利(Rayleigh,L.)
瑞梅基(Ramadge,P.)
瑞斯纳(Reissner,E.)

S

萨蒂(Saaty,T. L.)
萨方诺夫(Safanov)
萨弗罗诺夫(Сафронов,В. С.)
萨哈(Saha,M. N.)
萨林(Sarin,R. K.)
萨缪尔森(Samuelson,P. A.)
塞曼(Zeeman,E. C.)
塞曼(Zeeman,P.)
塞奇(Secchi,A.)
赛佛特(Seyfert,C. K.)
桑德尔(Sander,L. M.)
瑟斯敦(Thurston,E.)
沙可夫斯基(Sharkovsky,A. N.)
沙莱佛(Schlaiffer,R. O.)
沙利耶(Charlier,C. W. L.)
沙米尔(Shamir)

沙万奇(Savage,L. J.)
莎士比亚(Shakespeare,W.)
上田皖亮(Yoshisuke Ueda)
沈括(Shen Kuo)
生纳(Shanno,D. F.)
圣韦南(Saint-Venant,A. J. C. B. de)
施蒂费尔(Stiefel,C. F.)
施米特(Шмидт,О. Ю.)
施瓦布(Schwabe,S. H.)
施瓦茨席尔德(Schwarzschild,K.)
施瓦茨席尔德(Schwarzschild,M.)
施沃弗(Schwefel,H. P.)
舒尔(Schur,I.)
司马迁(Si Maqian)
斯卡里格尔(Scaliger)
斯里弗(Slipher,V. M.)
斯梅尔(Smale,S.)
斯盆得莱(Spendly,W.)
斯塔尔克(Stark,J.)
斯塔克伯格(Von Stackelberg)
斯特布勒(Stabler,R. C.)
斯特拉茨基(Strazicky,B.)
斯特龙根(Strömgren,B. G. D.)
斯特鲁维(Струве,В. Я.)
斯托克斯(Stokes,G. G.)
斯托梅尔(Störmer,F.)
斯瓦德士(Swadesh,M.)
宋德曼(Sundman,K. F.)
宋应星(Song Yingxing)
索曼(Solo,V.)
索米利亚纳(Somigliana,C.)
索绪尔(Saussure,F. De)

T

塔尔斯基(Tarski,A.)
塔克尔(Tucker,A. W.)
塔肯斯(Takens,F.)
太尔各特(Talcott,A.)
泰勒(Tayler,H. D.)
汤博(Tombaugh,C. W.)
汤超(Tang,C.)
汤姆森(Thomson,W.)
汤姆森(Thomson,J.)
汤普森(Thompson,K.)
特雷斯加(Tresca,H.)
特纳(Turner,H. H.)
梯塞朗(Tisserand,F.)
提丢斯(Titius,J. D.)
铁木森柯(Timoshenko,S. P.)
图灵(Turing,A. M.)
托勒密(Ptolemy)
托里切利(Torricelli,E.)
托姆(Thom,R.)

W

瓦尔德(Wald,A.)
瓦尔德迈尔(Waldmeier,M.)
瓦维洛夫(Вавилов,С. И.)
外尔(Weyl,(C. H.)H.)
王耿介(Wang Gengjie)
王浩(Wang Hao)
旺纳姆(Wonham,M.)
威顿(Witten,T. A.)
威尔(Will,C. M.)
威尔森(Wilson,E. L.)
威尔森(Wilson,R. W.)
威尔逊(Wilson,K. G.)
威尔逊(Wilson,R. H.)
威腾豪森(Witsenhausen,S.)
韦伯(Weber,J.)
韦茨(Wets,R.)
韦尔斯(Wells,A. A.)
维尔弗维茨(Wolfowitz,J.)
维尔兴(Wilsing,J.)
维默(Wymore,A. W.)
维纳(Wiener,N.)
维纳(Wiener,R.)
维宁・曼尼茨(Vening Meinesz,F. A.)
维森弗特(Wiesenfeld,K.)
维佐夫(Widrow,B.)
魏德利希(Weidlish,W.)
魏庆荣(Wei Qingrong)
魏札克(Weizsäcker,C. F. V.)
温伯格(Weinberg,S.)
文斯特雷曼(Vincent Rijmen)
沃尔夫(Wolf,J. R.)
沃尔夫(Wolf,J. E.)
沃尔夫(Wolfe,P.)
沃尔什(Walsh,D.)
沃尔什(Walsh,M. J.)
沃尔泰拉(Volterra,V.)
沃古瓦(Vauquois,B.)
沃克(Walker,A. G.)
沃勒(Wöhler,A.)
乌尔姆(Uim,H.)
吴文俊(Wu Wen-Chun)
吴有训(Wu Youxun)
伍兹(Woods,W.)

X

西尔亚克(Siljak,D. D)
西蒙(Simon,H. A.)
西姆森(Simson,R.)
西奈(Sinai,Ya. G.)
希尔(Hill,G. W.)
希尔伯特(Hilbert,D.)
希尔伏宁(Hirvonen,R. A.)
希契科克(Hitchcock,F. L.)

喜帕恰斯(Hipparchus,(R.))
夏普利(Shapley,L. S.)
夏禹(Xia Yu)
仙农(Shannon,C. E.)
谢多夫(Седов,Л. И.)
辛(Symplectic)
辛格(Singer,D.)
休伊什(Hewish,A.)
薛定谔(Schrödinger,E.)

Y

雅可比(Jacobi,C. G. J.)
雅库波维奇(Якубович,Б. А.)
亚当斯(Adams,J. B.)
亚当斯(Adams,J. C.)
亚当斯(Adams,W. S.)
亚里士多德(Aristotle)
延森(Jensen,J. L. W. V.)
央斯基(Jansky ,K. G.)
杨(Young,T.)
杨路(Yang Lu)
伊瓦赫年柯(Ивахненко,Г. А.)
依斯朵里(Isidori,A.)
英格维(Yingve,V.)
雍(Ljung,L.)
尤尔(Yule,G. U.)
尤里(Urey,H. C.)
於宗俦(Yu Zongchou)
约翰第三·伯努利(Bernoulli,Johann Ⅲ)
约翰逊(Johnson,M. A.)
约翰逊(Johnson,H. J.)
约翰逊(Johnson,H. T.)
约克(Yorke,J. A.)

Z

藻腾狄克(Zoutendijk,G.)
泽内尔(Zener,C.)
扎德(Zadeh,L. A.)
扎凯(Zakai,M.)
詹姆斯(Zames,G.)
詹森(Jensen,K.)
张德勒(Chandler,S. C.)
张景中(Zhang Jingzhong)
张遂(Zhang Sui)
郑和(Zheng He)
周江文(Zhou Jiangwen)
朱丁(Judin,D. B.)
朱利亚(Julia,G.)
朱思本(Zhu Siben)
朱斯(Joos,M.)
兹威基(Zwicky,F.)
祖绶(Zu Shou)

后　　记

十八载坎坷跋涉，千余人魂牵梦萦，这部涵盖现代数学科学体系的大型工具书——《数学辞海》终于杀青付梓了，释负之余感慨良多。

上世纪80年代中期，随着国家改革开放的深入，华夏盛世初显，我们这些数学工作者深感教学与科研急需，且人过中年应有所建树以无愧人生，于是决意编纂一部大型数学工具书，以振兴祖国数学事业，为中华民族争光。当《数学辞海》的选题一经提出，便在国内外数学界赢得热烈反响，特别是得到了前辈名家的亲切关怀和积极支持。又经广泛调研、民主磋商和反复论证，一部集古今中外数学成就于一体的《数学辞海》总体设计方案被确定下来，我们从此踏上了始料不及的艰难历程。

立意之初，我们考虑到国家百业待兴，财力紧缺，准备不靠国家拨款，自筹资金完成这项系统工程，闯一条民间编纂大型工具书的新路。为搞好编纂工作，特地组成了民间机构——数学辞海编辑委员会及其常设联络办事机构：数学辞海编辑部，并得到国家教育部、山西省教育厅、山西省新闻出版局和山西省教育学院（现与山西大学师范学院、太原师专合并为太原师范学院）等有关部门的认可。撰稿初期，由于有200余所院校及科研单位几代数学工作者的热情支持和积极参与，进展尚属顺利，但随着工程的进展，要在全国范围内（包括港、台地区）的1500多名专家、教授之间联系落实撰稿、统稿、审稿、改稿、编辑、校对等工作，再加上绝大多数的专家、教授是利用业余时间完成以上工作的，缺乏资金来源和专业的工作人员等困难，使之民间组织的数学辞海编辑部实在不堪重负。为解决编辑活动经费，编辑部的一些人几度成为当代"武训"，四处奔走，多方求助。就这样，编辑部仍经常处在邮资、通讯和差旅费难以支付的境地。

在经历了"九九八十一难"之后，在《数学辞海》终于诞生的今天，我们深深感谢社会各界及国内外有识之士给予的慷慨捐助，特别是山西省人民政府的资助；深深感谢山西教育出版社、东南大学出版社、中国科学技术出版社和北京大学出版社给予的关键性支持。我们也不能忘记那些给我们送来100元、500元、1000元……的捐助者，当然更要告诉读者的是：如果您感到此书对您稍有帮助的话，请不要忘记这1000多名数学工作者是不计报酬、不讲条件地编纂这部工具书的，他们当中还有很多人把自己的工资捐献给编辑部，以确保数学辞海编辑部的工作不致中断。还有一些专家、教授，历经数年，甚至十几年苦心修典，往往一天伏案十五六个小时，终于积劳成疾，竟然没有亲眼看到《数学辞海》面世，就不无遗憾地离开了我们。听着他们临终遗言："一定要尽快出版中国的《数学辞海》"，更增添了我们的一份紧迫感和责任感。

具有悠久历史的中华民族，对世界数学发展的杰出贡献，长期为世人瞩目，虽经中落，但中国当代数学科学又有了重大的进步。我们相信：在国家"科教兴国"方针指引下，中国必将再度成为数学大国，深望《数学辞海》能为实现这一宏伟目标略尽微薄之力。

《数学辞海》第一版即将面世之时，一种不名的恐惧萦绕心头，它的质量能获得读者的认可吗？能达到立意之初衷吗？希望广大读者在发现此书的种种问题时，不吝赐教。待我们稍稍喘息之后，将再邀请一批专家、教授对其进行修订，使之进一步充实提高，以期终臻完善。

数学辞海编辑部

2002年7月8日

《数学辞海》编辑部

图书在版编目(CIP)数据

数学辞海.第5卷/《数学辞海》编辑委员会编.
北京:中国科学技术出版社,2002.8
ISBN 7-5046-3330-5

Ⅰ.数… Ⅱ.数… Ⅲ.数学—词典
Ⅳ.O1-61

中国版本图书馆CIP数据核字(2002)第055359号

书　　　名:数学辞海(第五卷)
著作责任者:《数学辞海》编辑委员会
责 任 编 辑:理　石　胡　萍
书 名 题 字:启　功
装 帧 设 计:林胜利　王春声
标 准 书 号:ISBN 7-5046-3330-5/O·70
出　版　者:中国科学技术出版社
东 南 大 学 出 版 社
山 西 教 育 出 版 社
印　装　者:山西新华印业有限公司新华印刷分公司
发　行　者:中国科学技术出版社　东南大学出版社　山西教育出版社
经　销　者:新华书店
规　　　格:880mm×1230mm　1/16　63.375印张　2245千字
出 版 日 期:2002年8月第1版　2002年8月第1次印刷
本 卷 定 价:310.00元　(全套定价:1800.00元)